

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
19. April 2001 (19.04.2001)

PCT

(10) Internationale Veröffentlichungsnummer  
**WO 01/27760 A2**

(51) Internationale Patentklassifikation<sup>7</sup>: **G06F 11/00**

(21) Internationales Aktenzeichen: **PCT/EP00/10136**

(22) Internationales Anmeldedatum:  
13. Oktober 2000 (13.10.2000)

(25) Einreichungssprache: **Deutsch**

(26) Veröffentlichungssprache: **Deutsch**

(30) Angaben zur Priorität:  
199 51 209.4 14. Oktober 1999 (14.10.1999) **DE**

(71) Anmelder und

(72) Erfinder: **SCHLAFKE, Joachim [DE/DE];** Hauptstr. 63,  
53859 Niederkassel (DE).

(74) Anwälte: **HENGELHAUPT, Jürgen, D. usw.;** Gulde  
Hengelhaupt Ziebig, Schützenstrasse 15 - 17, 10117 Berlin  
(DE).

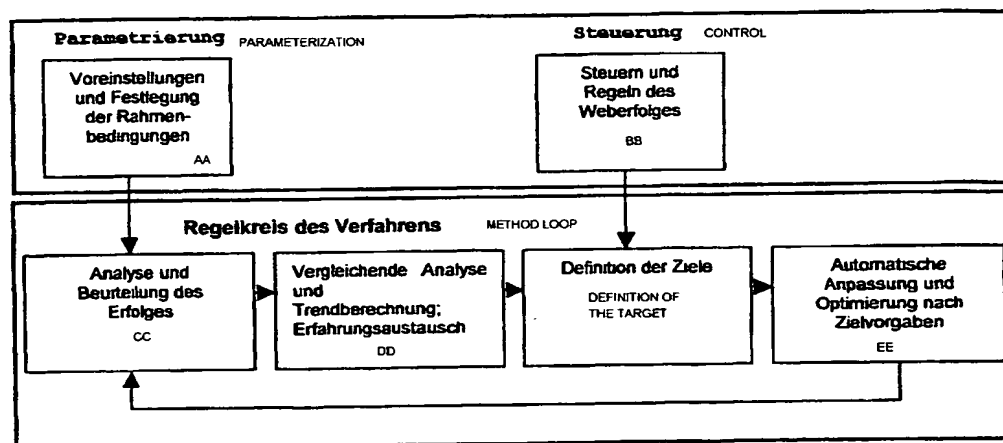
(81) Bestimmungsstaaten (*national*): AE, AG, AL, AM, AT,  
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Bestimmungsstaaten (*regional*): ARIPO-Patent (GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eura-  
sisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI,  
FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI-Patent

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR THE ANALYSIS OF USER BEHAVIOR IN A COMPUTER NETWORK IN ORDER TO OPTIMIZE  
WEB PRESENCE

(54) Bezeichnung: VERFAHREN ZUR ANALYSE DES BENUTZERVERHALTENS IN COMPUTERNETZEN ZUR OPTIMIE-  
RUNG DER WEB-PRÄSENZ



AA...PRESETTING AND DETERMINATION OF THE  
FRAMEWORK  
BB...CONTROL AND REGULATION OF THE  
WEBERFOLGES  
CC...ANALYSIS AND ASSESSMENT OF THE  
SUCCESS  
DD...COMPARATIVE ANALYSIS AND THE  
CALCULATION; EXCHANGE OF EXPERIENCE  
EE...AUTOMATIC ADAPTION AND OPTIMIZATION  
AFTER DEFINITION OF TARGET

(57) Abstract: The invention relates to a method for the analysis of user behavior in computer networks in order to optimize web presence. Said method can in particular be used to obtain qualified and future-orientated data on user behavior on the internet and intranet in order to influence one's own internet presence as a means for planning production and turnover. Said method provides a feedback circuit comprising an actual status analysis, designation of target functions, constant observation of user behavior, actual/desired comparison, performance of a trend calculation in addition to content and structure optimization of said web presence.

[Fortsetzung auf der nächsten Seite]

WO 01/27760 A2



(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.*

**Veröffentlicht:**

— *Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.*

---

**(57) Zusammenfassung:** Die Erfindung betrifft ein Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz, welches insbesondere eingesetzt werden kann, um qualifizierte und zukunftsorientierte Daten über das Userverhalten im Internet und Intranet zu erhalten, um damit Einfluss nehmen zu können auf die eigene Internet-Präsenz als Mittel zur Produktions- und Umsatzplanung. Das Verfahren stellt einen Regelkreis zur Verfügung, welcher besteht aus einer Ist-Analyse, der Vorgabe von Zielfunktionen, der ständigen Beobachtung des Userverhaltens, dem Ist/Soll-Vergleich, einer Durchführung der Trendberechnung und der Content- und Strukturoptimierung der Web-Präsenz.

---

**Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz**

---

5

Die Erfindung betrifft ein Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz, welches insbesondere eingesetzt werden kann, um qualifizierte und zukunftsorientierte Daten über das Userverhalten im Internet und Intranet zu erhalten, um damit Einfluß nehmen zu können auf die eigene Internet-Präsenz als Mittel zur Produktions- und Umsatzplanung.

15

Seit der Einführung des Internet (world wide web oder auch www) ist die Zahl der online-User weltweit enorm angestiegen. Durch diese Entwicklung ist das Internet zu einer wichtigen Plattform für die Wirtschaft auf internationaler Ebene geworden. Mit der fortschreitenden Kommerzialisierung wird das Netz immer multimedialer; Grafiken, Videos, Musik und Sprache ergänzen die reinen Textinformationen. Die ursprünglich auf reinen Informationsaustausch ausgerichtete Kommunikationskultur gestattet heute die Verbreitung von Werbebotschaften. Andere Medien wie Mobiltelefone oder TV-Geräte werden unmittelbar in den durch das Internet bereitgestellten Kommunikationsprozeß integriert. Somit ist Internet zu einem vollwertigen Kommunikationsinstrument im business-to-business (B2B) und ebenfalls im business-to-customer (B2C) Markt gereift. Dies führt zu einer drastischen Veränderung der Konsumgewohnheiten und des Konsumverhaltens der privaten Haushalte. Um den veränderten Anforderungen gerecht zu werden, ist es unter-

30

nehmensseitig unerlässlich, das Angebot an die veränderten Bedürfnisstrukturen der Konsumenten zu adaptieren. Innovative Unternehmen im Konsumgüter- und Dienstleistungsbereich erkennen das enorme Marktpotential und investieren in diese integrativen Kommunikationsplattformen. Um proaktiv die Zukunft dieses Mediums mitzugestalten, bedarf es detaillierter Informationen über die jeweiligen sich permanent modifizierenden User-Profile. Die Chancen durch ein Controlling-Instrument exakte User-Profile zu erstellen ist größer denn je, birgt aber auch die Gefahr der Fehlinterpretation der erhaltenen Daten.

Bei allen bekannten diesbezüglichen Verfahren wird nur quantitativ das Userverhalten protokolliert, basierend auf dem sog. Logfile, einem kontinuierlichen Logbuch, in welchen sämtliche Anfragen an den Server protokolliert werden. Ebenfalls, und das wird meist unterschlagen, wird eine Genauigkeit von den Systemen aufgezeigt, die bewiesenermaßen falsch ist. Das im Internet verwendet Verfahren des Proxy-Cash-Servers, welches eingesetzt wird um die Gesamtperformance des Internets zu erhöhen, führt beispielsweise zu keinem Eintrag im Logfile wenn die Seite über den Cash-Server geladen wird. In diesem Falle werden tatsächliche Besucher der entsprechenden Seite nicht registriert. Umgekehrt führt der Besuch von „ROBOTS“, also von Spionen der Suchmaschinen, zu Einträgen, die als Besucher der Website im Logfile protokolliert werden, aber im Sinne der Analyse keine sind. Sobald diese Situationen nicht berücksichtigt werden, ist die Interpretation mit Fehlern behaftet. Die Erklärung für diesen Zustand ist recht einfach. Der Logfile ist ein Format für die zeitliche Dokumentation des Webserver. Er ist nicht für die Analy-

se des Userverhaltens direkt konzipiert. Aus diesem Grund ist das Ergebnis des Logfile nicht falsch sondern wird vielmehr inkorrekt interpretiert.

5 Der Erfindung liegt deshalb die Aufgabe zugrunde, ein Verfahren zu entwickeln, mit welchem die Mängel der bekannten Verfahren vermieden werden, welches insbesondere

- von Fehlern bereinigte Daten zur Verfügung stellt,
- 10 - eine separierte Darstellung von Daten, die eine Fehlertoleranz aufweisen, inkl. der Darstellung der Bandbreite der Toleranz,
- eine Unterscheidung in Neu- und Altkunden sowie
- die Clusterung in Kundensegmente erlaubt,
- 15 - zukunftsorientierte Daten und Planungsdaten liefert, welche durch
- beliebig wählbare Analyse-Intervalle ermittelt werden können,

und darüber hinaus dem User eine erweiterte technische  
20 Performance wie z.B. Sprache, Bildschirmauflösung, Farbtiefe etc. zur Verfügung stellt und durch die historische Dokumentation des Internetauftritts dem User ein Kontrollinstrument für den Kommunikationsablauf in die Hand gibt, welches ihn in die Lage versetzt,  
25 schnell das Nutzerverhalten im Internet zu analysieren und darauf basierend erfolgreich Planungsmaßnahmen durchzuführen.

Diese Aufgabe wird erfindungsgemäß gelöst durch die  
30 Merkmale im kennzeichnenden Teil des Anspruchs 1 im Zusammenwirken mit den Merkmalen im Oberbegriff. Zweckmäßige Ausgestaltungen der Erfindung sind in den Unteransprüchen enthalten.

Ein besonderer Vorteil der Erfindung liegt darin, daß durch das Verfahren das Benutzerverhalten in Computernetzen analysiert und darauf basierend die eigene Web-Präsenz optimiert wird, indem nach Implementierung des Verfahrens der Quellcode der Domain optimiert wird, von mindestens einem Servercomputer Benutzeraktivitäten erfaßt und protokolliert werden, die erfaßten Aktivitäten von Interpretationsfehlern bereinigt werden, die nach der Bereinigung vorliegenden Daten analysiert und mit Daten aus einer Erfahrungsdatenbank verglichen werden und die Ergebnisse des Vergleiches zur Optimierung eingesetzt werden, indem durch eine Rückführung eine aktive Einflußnahme auf den Servercomputer erfolgt und hierdurch ein geschlossener Regelkreis realisiert wird.

Es wird eine Methode zur Optimierung der Web-Präsenz in Computernetzen zur Verfügung gestellt, durch welche das Benutzerverhalten erfasst und analysiert wird, wobei fehlerhafte Einflußfaktoren eliminiert werden und eine Optimierung der Web-Präsenz mit dem Ziel der Erhöhung der betriebswirtschaftlichen Wirksamkeit der Web-Präsenz erfolgt derart, daß aus den erfassten, bereinigten Benutzeraktivitäten und Aussagen aus einer Erfahrungsdatenbank ein Regelwert zur Realisierung eines geschlossenen Regelkreises gebildet wird, welcher Einfluß auf die Gestaltung der Web-Präsenz ausübt.

Die Erfindung soll nachstehend anhand von zumindest teilweise in den Figuren dargestellten Ausführungsbeispielen näher erläutert werden.

Es zeigen:

Fig. 1      Ablaufdiagramm des Verfahrens;

Fig. 2      Darstellung der am Verfahren beteiligten Komponenten;

Fig. 3 eine grafische Darstellung der Analysedaten.

Das erfindungsgemäße Verfahren erlaubt es einem Betreiber eines Servers für Internet-Domänen (Home-pages),  
5 alle Besuche, Aktivitäten und Umgebungsparameter der Besucher (User) dieser Domänen automatisch zu analysieren, diese Informationen mit anderen, in Datenbanken gespeicherten Daten zu vergleichen und auszuwerten und dadurch die Internetpräsenz seiner Domänen automatisch  
10 zu optimieren.

Das geht folgendermaßen vor sich:

Sämtliche Anfragen von Usern an den Server, auf dem die Homepage liegt, werden protokolliert. Diese Anfragen und die darauf folgenden Antworten des Servers werden  
15 im sogenannten Logfile zeitlich erfaßt. Diese Erfassung ist auf die technische Struktur des Servers ausgelegt, da die Serveraktivität protokolliert wird. Der Logfile wird von diversen herkömmlichen „Analyse“-Programmen als Basis für die Erfolgsmessung des Internets genutzt.  
20 Auf die Unrichtigkeit dieser Ergebnisse wurde bereits eingegangen. Bei der reinen Logfile-Analyse werden ausschließlich historische Daten erfaßt und dargestellt.

In den folgenden Abschnitten sollen die relevanten Vorgänge bei der Kommunikation zwischen Besucher und Server, auftretende Probleme und deren Lösung durch das Verfahren beschrieben werden.  
25

#### **HIT-Problematik:**

30 In der Anfangszeit des Controllings wurden einfach die Zeilen im Logfile gezählt. Jede Anfrage eines Besuchers löst einen sogenannten Hit aus. In der heutigen Zeit ist jedoch eine Hit-Zählung ungeeignet. Eine Bildschirmseite, die im HTML-Format kreiert worden ist,

kann neben dem Textkörper auch Multimediaelemente wie Bilder, Tondateien oder Video enthalten. Wird nun diese Seite angefordert, dann werden parallel die dazugehörigen Multimediaelemente geladen, die ebenfalls im  
5 Logfile protokolliert werden und einen Hit auslösen. Eine Seite mit vielen Multimediaelemente erzeugt also eine höhere Zugriffszahl und somit mehr Hits, als eine Seite, die rein auf Textbasis gestaltet ist.

In dem erfindungsgemäßen Verfahren wird der Hitzählung  
10 nur ein untergeordneter Status beigemessen, da dieser Wert nur einen Anhaltspunkt über den Trafic erlaubt.

In dem Verfahren werden deshalb weiterhin die Hits einer Domäne gezählt. Ihnen wird aber keine große Bedeutung beigemessen. Im Gegensatz zu den auf den gebräuch-  
15 lichen Analysemethoden beruhenden Zählungen werden die Hits allerdings genauer sein. Die genaueren Aussagen werden durch das temporäre Abschalten des Proxy-Cache-Servers und dem anschließenden Berechnen der nicht registrierten Zugriffe getroffen.

20

#### **Logfile-Problematik:**

Aufgrund der Verfälschungen durch fehlerhaft ausgewertete Hits kann also ein unkontrolliertes Erfassen der Hits nicht in Frage kommen. Es müssen also die Dinge  
25 erfaßt werden, die unabhängig von der grafischen Opulenz der Website sind. In diesem Fall bietet es sich an, die Anforderungen von einzelnen „Rahmendateien“ (HTML-Dokumente) zu zählen. Die Anforderung einer HTML-Seite wird in PageViews gezählt. Im Gegensatz zu  
30 PageViews werden bei dem vom Verfahren eingeführten PageImpression nur noch die Seiten gezählt, die als contentführend eingestuft werden.

Das Verfahren erlaubt die genaue Ermittlung von PageViews. Die Einheit "PageView" ist natürlich nicht re-



presentativ für eine Gesamtanalyse, gibt aber heute die Grundlage ab zur Berechnung des 1000er Kontaktpreises für Werbeflächen im Internet. Der Werbetreibende erhält durch diese Einheit den Eindruck, daß seine Werbebotschaft öfters angezeigt wurde, als es tatsächlich der Fall war.

Die Einheit "PageView" wird natürlich ebenfalls vom Verfahren gezählt. Hier treffen zwei Probleme aufeinander: zum einen werden zu viele Zugriffe gezählt, die das Verfahren in einem späteren Schritt korrigiert, zum anderen ruft die unten näher beschriebene Proxy-Cacheserver-Technik den gegenteiligen Effekt hervor: Seitenanforderungen werden aus Cache-Speichern bedient und führen zu keinem Eintrag im Logfile. Durch ein komplexes Verfahren wird jedoch sichergestellt, daß alle angeforderten HTML-Dokumente gezählt werden. Ein CGI-Image von nur wenigen BIT erzwingt einen Zugriff auf den Domäne-Server. Die eigentlichen Seiten können vom Cache-Server geladen werden. Mit dieser Methode bleibt eine gute Performance erhalten, es wird aber ein Hit auf dem Domäne-Server ausgelöst. Eine anschließende Berechnung ermittelt für diesen CGI-Impuls die nötige Anzahl von PageViews. In einem weiteren Schritt werden natürlich sämtliche Zugriffe von Suchmaschinen aus der Analyse der PageViews entfernt, denn diese sind keine Besucher im Sinne der Analyse. Als Ergebnis erhält der Betreiber einer Domäne eine 100% Aussage über die Anzahl der PageViews. Wie bereits erwähnt, ist dieser Wert nicht repräsentativ aber immer noch gebräuchlich in der Bestimmung der Werbeträgerleistung. Somit wird das Verfahren sämtliche Kennzahlen unterstützen und darüber hinaus genauer darstellen, auch wenn diese Werte keine oder nur geringe Aussagekraft haben.

**Frame-Problematik:**

Zwar wird heute PageView und PageImpression als Basis zur Berechnung des 1.000er Kontaktpreis angewendet, aber für ein effizientes Controlling sind diese Analysen immer noch nicht geeignet. Mit der Einführung der „Frame-Technologie“ besteht die Möglichkeit, eine Bildschirmseite aus mehreren HTML-Seiten aufzubauen. Beim Laden dieser Bildschirmseite wird jedoch jeder Frame mit einer dazugehörigen HTML-Datei und den ebenfalls verknüpften Bildern im Logfile protokolliert.

Je nach Frame-Struktur und grafische Opulenz einer Internetseite differiert der Quotient zwischen Anzahl von Zugriffen (Hit) und den gesehenen Bildschirmseiten (ContentView). Aus diesem Grund ist die Erfolgsmessung nach dieser Methode unzureichend.

Da sich im Laufe der Zeit die Programmierung in Frame-Technologie immer mehr durchgesetzt hat, ist die Lokalisierung der contentführenden Seiten zu Auswertungszwecken erstes Gebot und eins der entscheidenden Ziele des Verfahrens.

Natürlich entspricht jede erstellte Domäne bestimmten Erwartungshaltungen. Dies ist auch der Grund für die vielfältigen und unterschiedlichen Darstellungsformen von Internetseiten. Ein Seitenaufbau von 2 Frames oder ein Seitenaufbau von 7 Frames sind nichts ungewöhnliches und können sinnvoll sein. Das Verfahren berücksichtigt die individuellen Gestaltungsspielräume von Domäne-Betreibern. Das Verfahren ermöglicht das Auffinden von denjenigen Seiten, welche Definitionen über den Seitenaufbau beinhalten. Exakt diese Seiten beinhalten weder Navigation, werbeführende Botschaften noch Content-Seiten. Sie sind also ausschließlich für das technische Funktionieren der Seite verantwortlich und für die Beurteilung des Auftrittes uninteressant. Die-

ses Auffinden der „technischen“ HTML-Dokumente reduziert die Gesamtzahl der PageViews um ca. 30%. Die restlichen HTML-Dokumente werden Ihrer Bestimmung nach in einer Tabelle abgelegt und mit Attributen versehen, wie z. B. contentführend, Navigation oder werbeführend. Mit dieser Zuordnungs-Tabelle ist es möglich, aus dem Logfile alle HTML-Anforderungen ihrer Bestimmung nach korrekt zuzuordnen und zu zählen. In Kombination mit der schon oben beschriebenen Methode der CGI-Images ist so die exakte Anzahl von Zugriffen auf die einzelnen HTML-Dokumente sichergestellt. Somit ist erstmals die Ermittlung der ContentViews zu 100% möglich.

#### **CGI-Problematik:**

Eine weitere Herausforderung sind die dynamisch erzeugten Seiten. Hier wird aus Datenbanken entsprechend der Vorauswahl des Users die Seite serverseitig generiert und an den User zurückgeschickt. Aus diesem Grund wird kein Eintrag einer HTML-Seite im Logfile vorgenommen. Im Fall, daß dynamisch erzeugte Seiten aufgebaut werden, bzw. daß die Frame-Technik zum Einsatz kommt, sind alle bisher erläuterten Einheiten und Methoden für ein Controlling ungeeignet. Auf der anderen Seite sind gerade die so generierten Seiten für eine Produkt- und Finanzplanung hochinteressant, da sie in der Regel durch die Vorauswahl des Besuchers eine detaillierte Aussage über die Wünsche und Interessen des Besuchers bzw. Kunden enthalten.

Immer mehr HTML-Dokumente, gerade im e-commerce-Sektor, generieren Ihren Content dynamisch. D.h. in einem HTML-Rahmen werden Informationen aus einer Datenbank generiert und dargestellt. Aus diesem Grund ist es nicht direkt möglich aus dem Logfile einen Rückschluß zwischen dem aufgerufenen HTML Dokument und den dynamisch

generierten Informationen zu schließen. Das Verfahren ermöglicht aber neben dem Logfile ein weiteren eigen erzeugten File parallel zu erstellen, wo die Anforderungsparameter aus dem HTML-Dokument protokolliert werden. So ist es also möglich neben dem Rahmen HTML-Dokument auch die Informationsinhalte zu protokollieren und diese in einen Bezug zu setzen. Die Einheit ContentView wird, was auch den Tatsachen entspricht, um die Informationsinhalte von dynamisch erzeugten HTML-Dokumente quantitativ, also zahlenmäßig und qualitativ, also inhaltsbezogen, richtig erfaßt und dargestellt.

**Proxy-Cacheserver-Problematik:**

Proxy-Cacheserver dienen dazu, die vorhandene Bandbreite des Internets besser auszunutzen. Dieses wird durch den Einsatz von Cacheservern erreicht, indem diese Internetseiten zwischenspeichern. Fordert nun ein am Proxy-Server angeschlossener User eine Internetseite an, überprüft der Server ob die Seite bereits lokal vorliegt. Nur wenn die Seite nicht vorliegt, wird auf den Server des Seitenanbieters zugegriffen und ein Eintrag im Logfile vorgenommen. Dieser Eintrag enthält jedoch in den meisten Fällen die IP Adresse des Proxy-Cacheservers und nicht die des Users der die Seite angefordert hat. Im anderen Fall erhält der User die Seite aus dem Cacheserver, welches zur Folge hat, daß im Logfile des Seitenanbieters kein Eintrag erfolgt.

Zur genauen Auswertung ist es erforderlich, daß ein Seite die der User angefordert, serverseitig erfaßt und nicht unbemerkt und unregistriert vom Proxy-Cache an den User gesendet wird.

Die durch die Proxy-Cacheserver-Methode verbesserter Gesamt-Performance birgt auch einige Nachteile. Häufig aktualisierte Content-Seiten dürfen nicht in einem

Cache vorliegen, da sonst nicht sichergestellt werden kann, daß die darin enthaltenen Daten auch wirklich aktuell sind. Normalerweise wird in der Programmierung der Befehl „NOCACHE“, also „cachen nicht erlaubt“ eingegeben. Es gibt aber Provider, die diesen Befehl absichtlich unterdrücken, da ansonsten die Leistungsgrenze der Provider überschritten würde. Das Einfügen des Befehls „NOCACHE“ garantiert also nicht zu 100%, daß die Informationen direkt vom Domäne-Server abgerufen und so im Logfile protokolliert werden.

Dieser Problematik begegnet das Verfahren durch das Einbringen von erzwungenen Zugriffen auf den Domäne-Server bei content- bzw. werbe-führenden Seiten. Ein mittels CGI generierter Befehlssatz von nur 50 kb erzwingt bei jedem Browserzugriff auf eine Seite die Übertragung eines 1 Pixel großen Images vom Domäne-Server, auch wenn diese Seite bereits im lokalen Cache der Useres oder dem des Proxy-Servers vorliegt. Die Seite selbst unterliegt auf Wunsch weiterhin den Vorteilen des Cache-Mechanismus. Die Netzlast wird lediglich um jene 50 kb pro gezählter Seitenzugriff erhöht. Dieser Wert spielt in der heutigen Netzbandbreite keine Rolle und wirkt sich so nicht negativ auf die gesamte Performance der Seite aus.

25

**lokale Cachespeicher-Problematik:**

Ein vergleichbarer Effekt wird von dem lokalen Cachespeicher, der auf dem Rechner des Users vorliegt, ausgelöst. Alle bereits einmal übertragenen Seiten werden auf dem Rechner des Users im Cache-Verzeichnis abgespeichert. Dies führt zu einem beschleunigten Aufbau der Seite und wird heute von allen gängigen Browser unterstützt. Auch dieses Verfahren führt zu einer verrin-

30

gernden Zählung von Besuchen auf dem Server des Seitenanbieters.

Beispiel (Zeitverhalten): Ein User fordert die Seite 1 an und entschließt sich aufgrund eines Querverweises  
5 beim Lesen eine andere Seite aufzurufen. Nach kurzer Zeit wechselt er zurück auf Seite 1. Diese Seite wird nun aus dem lokalen Cache-Verzeichnis wiederhergestellt. Die Messung der Verweildauer im Logfile wird nun fälschlicherweise der Seite 2 zugeordnet unter der  
10 Prämisse, daß eine noch nicht im Cache vorhandene Seite vom Server angefordert wird. Hier ist zu beachten das die lokale Cachespeicher-Problematik mit der Proxy-Cacheserver-Problematik sich verstärkend auf die Fehlinterpretation auswirken kann.

15 Zur genauen Auswertung ist es erforderlich, daß ein Seite die der User angefordert, serverseitig erfaßt und nicht unbemerkt und unregistriert vom lokalen Browser-Cache an den User gesendet wird.

Die heute verwendeten Browser ermöglichen dem User, die  
20 Verwaltung des Cache-Speichers individuell einzustellen. Da es aus diesem Grund unzählige Variationen von Einstellungen gibt, kommt auch bei der lokalen Cache-Problematik das Verfahren welches bei der Diskussion der Proxy-Cacheserver-Problematik beschrieben  
25 ist, zum tragen. Ein mittels CGI generierter Befehlsatz von nur 50 kb erzwingt bei jedem Browserzugriff auf eine Seite die Übertragung eines 1 Pixel großen Images vom Domäne-Server, auch wenn diese Seite bereits im lokalen Cache der Useres vorliegt. Mit dem Verfahren  
30 wird neben der Proxy-Cache-Problematik auch die lokale Cache-Problematik gelöst.

**dynamische IP-Problematik:**

Eine weitere Herausforderung bei der Ermittlung von Besuchern stellt die Vergabe von dynamischen IP-Adressen dar. Viele Internet-Service-Anbieter vergeben sogenannte dynamische IP-Adressen. Diese IP-Adresse ist zwingend erforderlich in der Internet-Welt. Das Internet arbeitet nach dem Client-Server-Prinzip. Der Client stellt eine Anfrage an den Server. Der Server bearbeitet diese Anfrage und sendet das Ergebnis an die IP-Adresse der Anfrage zurück. Für die eindeutige Zuweisung der Nachricht ist also von beiden Seiten gesehen die IP-Adresse zwingend erforderlich. Die IP-Adresse ist in einem standardisierten Format beschrieben. (www.musterfirma.de oder 195.128.102.64). Es gibt nun weltweit mehr Internetnutzer als IP-Adressen. Aus diesem Grund müssen sich mehrere Internetnutzer wenige IP-Adressen teilen. Diese Methode müssen speziell große Dienstanbieter einsetzen. Hier bilden sich zwei Gruppen in der dynamischen IP-Vergabe.

Typ 1: Der User erhält bei der Einwahl in das Internet eine dynamisch IP-Adresse und die er für die gesamte Online-Session behält.

Typ 2: Hier erhält der User bei jeder Anforderung eines Bildschirmseiten-Elementes (Bild, Text, etc.) eine neue IP-Adresse. Bei starkem Datenverkehr beim Serviceanbieter wird bei einem erfolgreich beantworteten Element diese IP wieder freigegeben.

Sollten keine speziellen Zusatzverfahren angewendet werden, ist es also unmöglich die Anzahl der Besuche zu ermitteln. Auf dem letzten Internet-Kongreß wurde beschlossen, das Format um 1\*3 bzw. 2\*3 Stellen zu erweitern. Bis dieses weltweit in der Praxis eingeführt ist, werden wohl noch ca. 2 Jahre vergehen, in der die

Problematik durch den starken Zuwachs an Usern verstärkt wird.

Eine Bereinigung von möglichen Fehlinterpretationen von dynamischen IP-Adressen kann leider nur bedingt erreicht werden. Eine Möglichkeiten diese Problematik zu kompensieren, ist einen User unabhängig von seiner IP zu identifizieren. Ist dieser User nun festgestellt, dann kann er als einzelner Besucher erkannt werden, auch wenn er mit unterschiedlichen IP's innerhalb einer Session auftritt. Möchte der User sich nicht identifizieren, so besteht weiterhin die Möglichkeit über eine Zuordnungslogik die Wahrscheinlichkeit zu ermitteln, daß die unterschiedlichen IP's einem Besucher zugeordnet werden. Die Darstellung der Besucherzahlen erfolgt dann nach den Gesetzen der Aussagenwahrscheinlichkeit in Bandbreiten. Es wird also neben den tatsächlich vorhandenen Besucherzahlen eine weitere Grafik geben mit dem Höchstwert an Besuchern, dem niedrigsten Wert an Besuchern und der berechneten Anzahl von Besuchern.

Leider muß vorausgeschickt werden, daß die Problematik nicht 100%ig gelöst werden kann. Die Vorteile des Verfahrens beruhen jedoch auf der Erkenntnis, zwischen sicheren und unsicheren Daten zu unterscheiden, diese zu filtern und diese Daten in einer Bandbreite zu präsentieren. Durch das Verfahren kann diese Problematik sogar bis zu 98% Aussagegenauigkeit gelöst werden. Somit sind die Ergebnisse des Verfahrens so genau, daß normgerechte Planungsdaten berechnet werden können. 2% Abweichung vom realen Wert spielen für die strategische Weiterentwicklung keine entscheidende Rolle. Der Lösungsansatz beruht auf zwei grundsätzlich unterschiedlichen Userverhaltensmustern.

Fall 1: Der User identifiziert sich (wahlweise auch unter einem Nickname)



Fall 2: Der User möchte auf jeden Fall unerkant bleiben

Betrachten wir nun Fall 1: In diesem Fall hat der User  
5 die Cookies in seinem Browser aktiviert. In diesem Fall  
wird eine Variable in das Cookie eingetragen. Sollte  
nun der Besucher die Homepage besuchen, kann er während  
der Session anhand seines Cookieeintrages identifiziert  
werden. Besucht er nun zu einem späteren Zeitpunkt die  
10 Domäne erneut, so wird diesem Besucher ein erneuter Be-  
such zugewiesen, selbst wenn er unter einer neuen IP-  
Adresse erscheint. Mit dieser Methode (unter der Prä-  
misse das 100% der User Cookies aktiviert haben) kann  
eine eindeutige Aussage, die wie folgt aussehen könnte,  
15 getroffen werden.

„Im Monat September des Jahres 1999 hatten Sie auf  
Ihrer Homepage 650 Besuche. Diese 650 Besuche wur-  
den von 120 unterschiedlichen Besuchern ausgelöst.  
Sie haben im letzten Monat 40 neue Besucher gewon-  
20 nen. ... usw.“

Die Erfahrung zeigt aber, daß nur 50% der User Cookies  
aktiviert haben. In der Konsequenz bedeutet dieses nun,  
daß eine unbekannte Masse von 50% übrig bleibt, die  
25 nicht zweifelsfrei identifiziert werden kann. Eine Mög-  
lichkeit, diese Masse trotz unterschiedlicher IP's ein-  
zelnen bestimmten Besuchern zuzuordnen, ist die Verfol-  
gung, welche Seiten nacheinander aufgerufen werden, der  
sogenannte Clickstream. Hilfreich hierfür sind weitere  
30 Informationen, die über den normalen Logfile nicht pro-  
tokolliert werden. Ein weiterer Befehl für die Zuord-  
nung unterschiedlicher IP's auf einen Besucher ist  
„VIA“. Via enthält die gesamte Kette der Proxy-  
Weiterleitungen. Anhand der Weiterleitung, sprich des

Weges, und dem logischen Zusammenhang im Clickstream des Logfiles kann nach der Wahrscheinlichkeitsrechnung eine extrem hohe Aussagengenauigkeit getroffen werden. Die gesamte Sicherheit dieses Teilverfahrens beläuft  
5 sich auf mehr als 98% und ist daher als sehr sicher einzustufen.

**Router-Problematik:**

Viele Unternehmen setzen aus Sicherheitsgründen Firewall-Rechner ein. Ein Firewall-Rechner dient der Sicherheit im Unternehmen. Es wird ein „Nadelöhr“ geschaffen, durch das alle Informationen nach innen und nach außen fließen müssen. Dieses Nadelöhr ist wesentlich leichter zu überwachen als viele offene Schnittstellen nach außen. Dieser Firewall setzt sämtliche im  
15 Unternehmen verwendete IP auf eine einzige externe IP-Adresse um. Im Falle, daß dieses Unternehmen keine Proxy-Cache-Server verwendet, erscheint grundsätzlich nur diese Firewall-IP im Logfile. Das heißt, auch verschiedene User im Unternehmen erscheinen unter der  
20 gleichen IP-Adresse und machen eine Differenzierung ohne zusätzliche Verfahren unmöglich.

Beispiel: 3 Besucher eines Unternehmens besuchen  
25 gleichzeitig eine Internet-Domäne. Durch die Vergabe einer IP an alle drei User erscheint der Eindruck, daß eine Person sehr schnell ungeordnet durch viele Kapitel und Unterkapitel springt.

30

Verwendet das Unternehmen jedoch ein Proxy-Cache-System, welches ja empfehlenswert ist, ist die Situation ähnlich wie bei den großen Internet-Service-Anbietern.

Dies ist eine ähnliche Herausforderung wie die Lösung der dynamischen IP. Eine 100% Aussage kann nur getroffen werden, wenn der User sich identifizieren läßt. Sollte dies nicht der Fall sein, so kann nur nach statistischen Modellen ermittelt werden, wie hoch die Wahrscheinlichkeit ist, daß ein bestimmter User die Anfragen an den Server stellt. Untersuchungen und Tests, die durchgeführt wurden, zeigen, daß sich ca. 50% aller User identifizieren lassen, so daß eine hohen Aussagegenauigkeit über die Anzahl der restlichen Besucher getroffen werden kann. Mit diesen Lösungsansätzen kann eine bis zu 98% genaue Aussage über die Anzahl der Besucher getroffen werden.

Die Router- bzw. Firewall-Problematik läßt sich mit der gleichen Methode, wie im Falle dynamischer IP's beschrieben, lösen (Clickstream, Verfolgen der Proxy-Weiterleitungen). Wobei jedoch anzumerken ist, daß die Aussagengenauigkeit beim Auftreten dieser Problematik etwas niedriger ist. Sie liegt bei ca. 92%. Dieser Wert fließt jedoch nur zu einem geringen Maße und branchenabhängig in die gesamte Auswertung ein und sollte individuell je nach dem Informationsbedürfnis berücksichtigt werden.

Eine weitere Möglichkeit zur Identifizierung kann, unter der Voraussetzung das PHP auf dem Server installiert ist, durch „Sessiontracking“ erreicht werden. Alle Dokumenttypen werden dann von \*.htm bzw. \*.html in \*.php umbenannt. Beim Aufruf der Startseite der Domain (meist Index.php) wird eine eindeutige Variable dem User zugeteilt. Diese Variable bleibt dem über die ganze Session konstant und wird an alle aufgerufenen Dokumente der Session angehängen. Diese Anhang ist im Standard- Logfile sichtbar und für die Auswertung selektierbar. Sessiontracking ist mit Sicherheit der „Königsweg“. In

Kombination mit Aktivierung der Cookies oder freigeschaltete Pentium III ID ist eine nahezu 100% Identifizierung des Userverhaltens möglich. Ebenfalls kann zweifelsfrei festgestellt werden, wie oft ein User die  
5 Seite besucht hat. (Differenzierung Neu-/Altkunde)

**Robot-Problematik:**

In regelmäßigen Abständen besuchen Suchmaschinen die Internetseiten, um nach Neuerungen und Aktualisierungen  
10 zu suchen. Diese Besuche geschehen vollautomatisch durch sogenannte Suchrobots. Suchrobots sind die kleine Softwarespione, die das Netz absuchen und sämtliche Neuerungen in das Datenbanksystem des Suchmaschinenbetreibers schreiben. Diese Besuche lösen natürlich  
15 ebenfalls Zugriffe im Serverlog aus, die fälschlicherweise als Besuch gewertet werden könnten. Diese Visits sind mit Sicherheit am einfachsten heraus zu filtern, sollten aber nicht unbeachtet bleiben.

Ziel muß es sein, die Besuche von Suchmaschinen auf einer Domäne aus dem Logile zu filtern, so daß diese  
20 nicht die Auswertungsdaten verfälschen. Weiterhin sollen diese Besuche in einer separaten Auswertung dargestellt werden, um Informationen über die verwendeten Suchmaschinen mit den dazugehörigen Suchworten zu erhalten.  
25

Fast alle gängigen Suchmaschinen greifen auf die Datei „ROBOT.TXT“ zu. Ein User wird niemals auf diese Datei in der Verzeichnisstruktur zugreifen. Aus diesem Grund kann im Logfile der Besuch von Suchmaschinen gefiltert  
30 werden und von den Userdaten abgezogen werden. Der Besuch von Suchmaschinen kann nun in eine separate Auswertung dargestellt werden. Aus dieser Auswertung können nun weitere Entwicklungsstufen für einen optimierten Metatag eingeleitet werden.

**Suchmaschinen- und Link-Problematik bei frame-orientierten Seiten:**

Diese Problematik wird durch mangelndes Wissen der Seitenhersteller über die Internetfunktionsweise ausgelöst. Wenn eine Suchbegriff in eine Suchmaschine eingegeben wird, und die Suche erfolgreich war, werden die Ergebnisse der Suche dem Anfragenden in der Ergebnisliste angezeigt. Klickt dieser nun auf die angegebene URL, dann erscheint meistens nur die Contentseite ohne die dazugehörige Framestruktur.

Leider findet man auf den meisten Seiten, die frameorientiert aufgebaut sind, diesen Fehler. Verstärkt wird dieser Fehler durch den Komfort und die Leistungsfähigkeit moderner Suchmaschinen. Wenn eine URL bei einer Suchmaschinen angemeldet wird, dann werden die Suchroboter diese URL innerhalb der nächsten Tage besuchen. Mit verschiedenen Parametern mit METATAGS beziehungsweise Parametern im ROBOT.TXT wird es zugelassen bzw. verhindert, daß bestimmte Unterverzeichnisse durchsucht werden. Wird das Durchsuchen von Unterverzeichnissen zugelassen, was ja auch sinnvoll ist, muß dafür Sorge getragen werden, daß die komplette Framestruktur im Falle eines Userzugriffes über die Suchmaschine mit aufgebaut wird, um die Navigation zu ermöglichen. Nur so ist es möglich die komplette Internetdomäne dem User zu präsentieren. Diese Programmierung wird meistens vergessen.

Aus einschlägigen Untersuchungen ist bekannt, daß 40% aller Ergebnisse von Suchmaschinen in eine Unterverzweigung der Homepage führen. Diese große Zahl von „Quereinsteigern“ muß ordnungsgemäß bedient werden. Eine programmiertechnische Überarbeitung, gerade der von Laien erstellter Homepages ist unumgänglich, da diese

Seiten auch noch x-fach in der Ergebnisliste der Suchmaschine angezeigt werden.

Ziel muß es sein, daß beim direkten Aufruf einer Content-Seite die gesamte Frame-Struktur aufgebaut und der User zu dieser Seite durchgeschleift wird.

Die Erfahrung zeigt, daß bei einer nicht unerheblichen Anzahl von Domänen gravierende Fehler im Quellcode vorliegen, bzw. Metatags falsch oder gar nicht generiert worden sind. Ein weiterer potentiell vorliegender Fehler, gerade bei frame-orientierten Internetstrukturen, ist der falsche Aufbau der Seite, wenn sie in einem untergeordneten Content aufgerufen wird und nicht von der Homepage aus. In diesem Fall wird die Navigation nicht mehr angezeigt. Diesen Problematiken wird bei dem Verfahren mit den geeigneten Methoden begegnet, die automatisch bzw. halbautomatisch den Code auf geeigneter Weise optimieren und korrigieren.

Die oben beschriebene Maßnahme ist für das Gesamtverfahren von großer Bedeutung. Zum ersten wird der Qualitätssicherung genüge getragen und zum zweiten liegen die Daten des Metatags und somit die Informationen für die Suchmaschine in geeigneter Form in einer Datenbank vor. So ist es erstmals möglich einen Soll/Ist-Vergleich zwischen den meisten gesuchten Schlagworten, den Suchmaschinen Fabrikaten und den tatsächlich vorliegenden Informationen auf einer Internetseite durchzuführen. Die Ergebnisse dieser Vergleichsstudie ermöglichen die gezielte Optimierung der Metatags für kundenorientierte Suchergebnisse und somit für den Erfolg einer Domäne.

30

#### **Interpretations-Problematik:**

Die einzige Möglichkeit Daten über das Userverhalten (es sei denn, der User identifiziert sich) zu erhalten, ist der Logfile. Der Logfile ist aber in seiner Funkti-

on für die Protokollierung der Servertätigkeit ausgerichtet und nicht unbedingt für weitere Interpretationen bestimmt. Einige Hersteller von Interpretationssoftware lesen die Daten aus und geben diese ungefiltert und nicht an die Bedürfnisse angepaßt an diverse Analysetools weiter. Fehlende Informationen über das Userverhalten und die technische Performance werden einfach angenommen bzw. übernommen. Die oben genannten Problematiken bleiben zum großen Teil unberücksichtigt. Die Konsequenz sind pseudogene Basisdaten, die weder für das Controlling noch für die Planung akzeptabel sind. Im Gegenteil: Es werden sogar falsche Daten als Bemessungsgrundlage aufgeführt, die eine Planung in die falsche Richtung verursachen und somit Schaden anrichten.

Ziel ist es, daß der Seitenanbieter zu Planungs- und Beurteilungszwecken nur die Daten erhält, die valide und aussagekräftig sind. Die Daten, die ausschließlich auf Vermutung basieren, sind in der Analyse zu eliminieren, so daß diese die Auswertung nicht verfälschen können.

Beispiel: Einige Analyse-Programme treffen die Aussage, daß 99% der User Cookies verwenden. Diese Aussage basiert auf dem Prozentsatz der cookiefähigen Browser. In Wirklichkeit haben ca. 50% bis 60% der User „Cookies akzeptieren“ ausgeschaltet, so daß der Wert „99% der User haben Cookies“ falsch ist. Solche Zahlen können zu dramatischen Fehlentscheidungen im e-commerce-Sektor führen. Ein Unternehmen, welches sein Shop-System cookiebasiert aufbaut, beschneidet seinen Kundenkreis aufgrund der Falschaussagen um 50%-60%. Sollte dieses bereits in den Brunnen gefallene Kind zu spät

erkannt werden, sind Umsatzeinbußen zu ver-  
zeichnen, die mit einem normgerechten Tool  
nicht entstanden wären. Eine Neuprogrammierung des Shopsystems unter Java, CGI, etc.  
5 ist unumgänglich.

Unsichere Daten bzw. empirisch ermittelte Daten sind  
auf Ihre Aussagengenauigkeit und Wahrscheinlichkeit zu  
prüfen und in den daraus resultierenden Toleranz-  
10 bereichen darzustellen.

Die hier dargestellten Herausforderungen werden durch  
die einzelnen Module des Verfahrens kompensiert. Die  
Erfinder des Verfahrens maßen sich nicht an, alle Daten  
mit einer 100% Sicherheit zu erhalten. Bei diesem Ver-  
15 fahren wird aber unterschieden zwischen Aussagen, die  
100% sicher und geprüft sind, und den Daten, die mit  
einer Ungenauigkeit behaftet sind. Diese unsicheren Da-  
ten werden nach anerkannten statistischen Methoden auf  
Ihre Aussagewahrscheinlichkeit geprüft und in einem To-  
20 leranzfeld dargestellt.

Der User kann zwischen drei Varianten der Darstellung  
wählen.

1. Darstellen von den Daten, die sicher sind, und ge-  
trennt davon
- 25 2. Darstellung der Daten, die in einem Toleranzfeld  
liegen, sowie
3. die gemischte und gewichtete Darstellung beider  
Daten mit ebenfalls enthaltenen Toleranzfeld.

Auf diese Weise kann sichergestellt werden, daß keine  
30 Daten fehlerhaft interpretiert werden und so Fehl-  
entscheidungen weitestgehend ausgeschlossen werden kön-  
nen.

**Format-Problematik:**



Zur Zeit sind mehr als 30 verschiedene Logfile-Formate im Einsatz. Je nach Betriebssystem, Serverarchitektur und Provider kommen nicht nur die 30 verschiedene Basis-Formate vor, sondern auch noch die unterschiedlichsten Kombinationen. Eine weitere Problematik ist die Verwaltung dieser Daten. Viele Provider löschen historische Daten oder komprimieren diese und legen sie in entsprechenden Unterverzeichnissen ab. In diesem Fall ist natürlich das Wissen des Internet-Administrators über den Umgang mit den Logfiles von großer Bedeutung. Von unabhängiger Entscheidungsfindung der Geschäftsleitung kann keine Rede mehr sein. Diese Vielzahl an Basis-Formaten des Logfiles, unterschiedlichen Kombinationen und eine uneinheitliche Verwaltung der historischen Daten auf Providerseite machen bisher ein einheitliche, lückenlose und vergleichbare Performance-Messung der Internet-Aktivitäten nahezu unmöglich. Das Ziel, die Format-Problematik zu kompensieren, kann nur erreicht werden, wenn die Daten in einen eigenen und unabhängigen Datenbanksystem protokolliert werden. Im besonderen ist zu berücksichtigen, daß herkömmlicherweise viele Provider die historischen Daten am Monatsende mit dem neuen Monat überschreiben. Eine langfristige Planung und freigestaltbare Analysezeiträume sind somit nur in den seltensten Fällen möglich.

Von der Basis her werden durch das Verfahren die unterschiedlichen Formate komplett unterstützt. Zu Beginn des Einsatzes des Verfahrens wird unter anderem ein Logfile downloadet und analysiert. Aus dieser Analyse wird der Informationsgehalt ermittelt und mit dem Anforderungsprofil des Controllers verglichen. Spezielle Zusatzinformationen wie z.B. Bildschirmauflösungen, „Cookies aktiv/nicht aktiv“ und die Ermittlung von Neu-/Altkunde sowie die Beseitigung von Proxy-Cache-

Problematiken werden in einer Erweiterung des Logfiles abgespeichert. Die wohl größte Herausforderung, die Beseitigung der unterschiedlichen Handhabung von Logfiles auf Providerseite wird durch das automatische Downloaden des Logfiles behoben. Die bereinigten Ergebnisse werden anschließend in ein eigens erstelltes Datenbanksystem lokal auf dem Rechner des Controllers abgespeichert. Aus diesem Datenmanagement-System werden alle weiteren Darstellungen generiert. Mit dieser Methode ist es möglich, verschiedene Zeitintervalle für die Analyse auszuwählen, bzw. verschiedene Zeitintervalle gegenüberzustellen.

Dieses Verfahren bietet somit einige entscheidende Vorteile:

- Die Datenbank ist unabhängig vom Logfile;
- Die Datenbank kann logisch mit anderen Datenbanken verknüpft werden (Soll/Ist-Vergleich);
- Es können individuelle Zusatzinformationen (z.B. Aktualisierungen) gespeichert werden;
- Unabhängigkeit vom Handling des Providers ist gesichert;
- Schnittstellen zu anderen Applikationen (z.B. zu Warenwirtschaftssystemen) werden bereitgestellt,
- Beliebige Bestimmung der Analysezeiträume wird gestattet;
- Es stellt eine Basis dar für Erweiterungen und Flexibilität bei nachfolgenden Ausbaustufen (mail-flow, comm-flow, etc.).

### **Darstellungs-Problematik:**

Im Logfile werden alle Protokolldaten chronologisch abgespeichert. Um diese Daten „lesbar“ zu gestalten, müssen diese gefiltert, sortiert und interpretiert werden. Dieser Aufwand wird selten oder nur mit hohen Kosten

betreiben. Es kann nicht ohne weiteres ein Rückschluß von der technischen Verzeichnisstruktur, die im Logfile erkennbar ist, und der Präsentationsstruktur gezogen werden. Ebenfalls werden alle Dateien, auch die für die  
5 Navigation bestimmten Daten, unstrukturiert dargestellt. Eine weitere Herausforderung ist es, einen Bezug zwischen den Dateinamen und den präsentierten Inhalten zu finden.

Die Folge ist: Ausschließlich der Personenkreis, der  
10 die Internetseite entwickelt hat, ist in der Lage den Logfile und die Auswertung zu interpretieren. Es ist aber ein großer Aufwand für die Interpretation der Daten erforderlich, der in den seltensten Fällen betrieben wird.

15 Die Vertriebs- und Marketingabteilung hat keine Möglichkeit, unabhängig von der Entwicklungsabteilung der Internetseite, strategische Entscheidungen über die Zukunft der Domäne zu treffen.

Die Darstellung muß deshalb in der Form aufgearbeitet  
20 sein, daß ein Personenkreis ohne herausragende Kenntnisse im Internet, die Ergebnisse verstehen kann. Die Basis ist die klare Strukturierung des Contents, die unabhängig von der technischen Strukturierung ist. Ebenso ist die Darstellungsform "Text" als Träger der  
25 Information nicht zu unterschätzen und sollte neben der grafischen Darstellung ergänzend wirken. Frei gestaltbare Bemessungszeiträume und eine individuelle Reduzierung der Daten auf das Informationsbedürfnis des Interessierten werden deshalb vom Verfahren zur Verfügung  
30 gestellt.

Voraussetzung einer leichtverständlichen Darstellung ist die exakte Abbildung der Darstellungs-Struktur einer Internetseite. D.h. die Struktur muß dem entsprechen, was sich dem User auf dem Monitor zeigt. Diese

Struktur entspricht in den seltensten Fällen der technischen Struktur auf dem Server. Aus diesem Grund ist es zwingend erforderlich, eine „Übersetzung“ zwischen der technischen und der Darstellungsstruktur zu schaffen.

5 Durch diese „Übersetzung“ werden auch die Bezeichnungen der Dateien in eine für den Laien verständlichen Sprache gebracht. Wie bereits erwähnt steht zum Anfang des Verfahrens das Downloaden der kompletten Internet-Struktur als Offline-Version. Das Verfahren filtert nun alle

10 Bilder aus, da diese für das Userverhalten unerheblich sind. Ebenfalls unerheblich sind alle HTML-Seiten, die den Rahmen für den Content bilden, d.h. einen Frameset aufrufen. Die nun übriggebliebenen Dateien entsprechen nur noch folgenden Kriterien: Navigation, Content, Wer-

15 beflächen, Ausführende Seite, Download.

Die Dateien werden nun vom Controller den Punkten Navigation, Content und Werbefläche zugeordnet, indem Sie nacheinander angezeigt werden. Nach dieser Prozedur stehen die contentführenden Seiten fest und werden in

20 Kapitel und Unterkapitel strukturiert. Diese Informationen werden ebenfalls in dem Datenbanksystem abgespeichert und dient als Basis für die Auswertung und Darstellung der Analyse.

Ein weiterführende Abfragesystematik analysiert die individuellen Informationsbedürfnisse des Controllers.

25 Diese Daten werden ebenfalls im Datenbanksystem gespeichert. Mit diesem Verfahren wird ein starker Visualisierungsgrad erreicht, der den geforderten Ansprüchen dienen und das wesentliche erfaßt.

30 Das Verfahren umfaßt die klassische Logfile-Analyse zur Dokumentation der chronologischen Abläufe auf Serverseite, wie die Dokumentation der Hits sowie der Errors der einzelnen HTML-Seiten im zeitlichen Kontext, die

Darstellung von diversen Seitenzugriffsstatistiken in Form von Balken- und Kreisdiagrammen und die Ermittlung des vom Besucher benutzten Browser-Typs. Erweitert wird die klassische Analyse durch ein Web-Analyse-System für  
5 eine aussagekräftige Dokumentation des tatsächlichen User-Verhaltens im Internet. Dieses Web-Analyse-System umfaßt die programmtechnische Optimierung des Internet-Auftrittes, die Darstellung fehlerbereinigter Informationen über das User-Verhalten, wie z. B.: content  
10 views, erfolgte Downloads, Pageviews; die Werbeträgerleistung nach IVW-Methode, die Darstellung des exakten zeitlichen User-Verhaltens, Informationen über den User wie verwendetes Betriebssystem, Browser, Farbtiefe, Auflösung des Monitors, tatsächliche Cookie-  
15 Aktivierung, JAVA-Fähigkeit, VB-Fähigkeit; Darstellung der Traceability des Users: verwendete Suchmaschinen und Suchworte, Herkunft des Users über Links, Direkteinwahl oder Einwahl über Subdomain, Analyse der typischen Userpfade: Ein- und Ausstiegsseite, Verweil-  
20 zeiten, Alt-/Neukunde; die Analyse der Verweilzeiten sowie ein Alarmsystem zur Analyse gering frequentierter Seiten. Ein weiterer Vorteil des Verfahrens gegenüber herkömmlichen Verfahren ist darin zu sehen, daß mit dem erfindungsgemäßen Verfahren ein Tool zur Analyse belie-  
25 biger Zeiträume unter Berücksichtigung des relativen und des absoluten Erfolges zur Verfügung steht. Durch den Zugang zu einer externen Erfahrungsdatenbank, in der das User-Verhalten dokumentiert wird, sind aussagekräftige Trendanalysen zukünftiger Entwicklungen und  
30 eine Evaluierung der User-Akzeptanz möglich.

Ein weiterer Teil des Verfahrens ist das Tool zur Optimierung bestehender Web Sites, welches automatisch eine

Code-Optimierung und Frame-Korrektur sowie Suchwort- und Meta Tag-Optimierung durchführt. Ebenso ist ein Tool für die Aktualisierung der Suchmaschinen-Einträge und der Eintragung selbst vorhanden. Ein weiterer Bestandteil wird in einem Trainings-Tool zur Verfügung gestellt. Dieser erlaubt ein web-basiertes Training, die Analyse zur Lernerfolgsmessung und die Evaluierung der Effizienz des Schulungsprogrammes. Ein Schulungsprogramm für effektives Web-Controlling unter Berücksichtigung rechtlicher Aspekte und technischer Voraussetzungen vermittelt die Darstellung der Basics sowie der Standardmethoden und bietet ein strategisch aufgebautes Controllingsystem zur Optimierung der interaktiven Kommunikationskette. Ergänzt wird das Verfahren durch eine Erweiterung um die typischen Funktionalitäten des e-commerce (PAP-Programmierung, CGI-Programmierung)

Die eben beschriebene Funktionalität des Verfahrens wird ergänzt durch die Möglichkeit, basierend auf der Datenanalyse Zielvorgaben-Kontrollfunktionen und Alarmsysteme zu etablieren, die ein effizientes Controlling ermöglichen, denn diese Daten sind für die strategische Weiterentwicklung und Optimierung der Internet-Domain unter dem Aspekt der Kundenorientierung von höchster Relevanz. Es wird damit ein Link zwischen der Analyse und der Optimierung geschaffen, um einen kontinuierlichen Verbesserungsprozeß (KVP) zu realisieren. Weitere Funktionalitäten, die vom Verfahren bereitgestellt werden, bestehen in der Verifizierung der Werbewirksamkeit von Marketingmaßnahmen durch spezielle Analyse der Web Site, der Verifizierung der Response-Aktivitäten des Unternehmens auf Anfragen über die In-

ternet-Domain und einer Beurteilung der kommunikativen Aktivitäten auch außerhalb des Webs.

Das Verfahren in seiner Gesamtheit stellt somit einen  
5 Regelkreis dar, der am Beispiel eines Internet-  
Auftritts mit einer Homepage wie folgt beschrieben werden kann. Ausgehend von der Präsentation der Internet-  
Seite erfolgt im Anschluß die Erfassung und Analyse der  
Benutzeraktivitäten, die von Erfassungs- und Interpre-  
10 tationsfehlern bereinigt werden. Über den anschließenden Vergleich mit von einer Erfahrungsdatenbank zur  
Verfügung gestellten Informationen erfolgt  
dann die Optimierung und Anpassung der Internet-Seite,  
mit dem Ziel die Vorgaben der Zielvorgaben-  
15 Kontrollfunktionen zu erreichen.

Detailliert kann der Prozeß folgendermaßen beschrieben werden:

Sämtliche Anfragen von Usern an einen Server werden  
20 protokolliert. Diese Anfragen und die darauf folgenden  
Antworten des Servers werden im sogenannten Logfile  
zeitlich erfaßt. Diese Erfassung ist auf die technische  
Struktur des Servers ausgelegt, da die Serveraktivität  
protokolliert wird. Der Logfile wird von diversen Pro-  
25 grammen als Basis für die Erfolgsmessung des Internets  
genutzt. Auf die Unrichtigkeit dieser Ergebnisse wurde  
bereits ausführlich eingegangen. Bei der reinen Log-  
file-Analyse werden ausschließlich historische Daten  
erfaßt und dargestellt. Zukunftsorientierte Informatio-  
30 nen können nicht berechnet werden, da bereits die Ein-  
gangsinformationen mit Fehlern behaftet sind.

Folgende Informationen werden im klassischen Logfile  
dargestellt :

- Hits:** Darstellung aller Zugriffe incl. Laden von Bildern und Aufbau der Frame-orientierten Seiten;
- PageViews:** Darstellung aller HTML-Seiten - das Laden von Bildern wird nicht mitgezählt. Dieser Wert  
5 wird heute als Größe zur Berechnung der Werbeträgerleistung herangezogen;
- Errors:** Fehler, wenn Seiten angefragt werden, die nicht auf dem Server vorhanden sind bzw. nicht gefunden werden können;  
10
- Übertragene Bytes:** Menge von Daten, die vom Server übertragen worden sind;
- Tagesverhalten:** Zeitliche Darstellung der Informationsanfragen; Angabe der Uhrzeit, wann Informationen abgefragt wurden  
15
- Wochenverhalten:** Zeitliche Darstellung der Informationsanfragen in der Tagesübersicht;
- Monatsverhalten:** Zeitliche Darstellung der Informationsanfragen in der Wochenübersicht;
- 20 **Allgemeines zeitliches Verhalten:** Verweildauer auf einer Seite, mit dem Hintergrund der Intensitätsprüfung und Prüfung der Kommunikationsstärke der Seite;
- Betriebssystem:** Zahlenmäßige Erfassung, mit welchem Betriebssystem der User arbeitet;  
25
- Browser:** Zahlenmäßige Erfassung, mit welchem Browser der User arbeitet;
- Suchmaschinen:** Statistische Erhebung, mit welcher Suchmaschine die Domain gefunden wurde;
- 30 **Suchworte:** Statistische Erhebung, mit welchen Suchworten die Domain gefunden wurde;
- Herkunft des Users:** Erfassung des Landes, aus dem der Besucher der Domain kommt;



**User-Pfade:** Erfassung der Klickfolge innerhalb der Domain;

**Ein-/Ausstiegsseite:** Erfassung der Ein-/Ausstiegsseite Users innerhalb einer Domain

5

Wie bereits dargelegt, reichen die o.g. Daten zur Realisierung eines effizienten Controllings jedoch nicht aus. Basierend auf der klassischen Logfile-Analyse bietet das Verfahren folgenden Zusatznutzen:

10

- Transport der technisch orientierten Information in die reale Darstellungsform der Domain
- Kontrolle, ob die Daten richtig erfaßt wurden
- Bereinigung von technisch bedingten Fehlinterpretationen
- Empirische Ermittlung von unsicheren Informationen mit der Darstellung in Toleranz-werten
- Darstellung der Daten entsprechend der Bedürfnisstruktur des Controllers

20

Die Bedürfnisstruktur läßt sich lt. neuesten Untersuchungen wie folgt darstellen :

25

- Was ist unter zeitlichen Gesichtspunkten auf der Domain geschehen?
- Wie ist es um die Leistungsfähigkeit der Rechner-Performance der User gestellt?
- Wie hat der User die Domain gefunden, wonach hat er gesucht und woher kommt er?
- Was war die Intention des Users, hat er die Domain bereits vorher besucht und wie lange hat er welche Informationen abgefragt?
- Administrative Zusatzinformationen zur Überprüfung der Kommunikation. Wurden alle Mails beantwortet,

30

sind die Zielvorgaben erreicht worden, welchen Einfluß haben Anzeigen und Werbung z.B. in Printmedien auf die Zugriffe der Domain etc.

- 5 Durch das Verfahren können weitere für den Controller interessante Informationen generiert werden. Hier ist eine klare Erweiterung der Leistungsfähigkeit gegenüber den bisher bekannten Produkten zu sehen. Und zwar zeichnet sich das R.T.F.I. Verfahren durch folgende Zu-
- 10 satzfunktionen aus :

**ContentViews:** Anzahl von Seitenkontakten, die als contentführend bezeichnet werden können, also tatsächlich relevanten Inhalt haben, und nicht der

15 Navigation dienen. Ein weiterer sekundärer Nebeneffekt ist die Analyse von Bannerwerbung und die daraus resultierende Abrechnung der Werbekosten. Diese werden zur Zeit aus den nicht korrekt erfaßten PageViews ermittelt;

20 **Seiten Statistik:** Wie verteilt sich die Besucherhäufigkeit auf den einzelnen Seiten zum gesamten Content? Wie sieht die Seite aus und wie lange wurde sie im Durchschnitt besucht?

**Downloads:** Welche und wie oft wurden bestimmte Dateien

25 downloadet?

**Infoanforderung:** Wie oft wurde die Seite mit einer Infoanforderung besucht und wie oft wurde anschließend die Information tatsächlich angefordert?

30 **Aktualisierung:** Wann und was wurde an einer bestimmten Seite verändert bzw. aktualisiert?

**Farbtiefe:** Welche Farbtiefe hat der User auf seinem Rechner eingestellt?

- Auflösung:** Welche Auflösung hat der User auf seinem Rechner eingestellt?
- Prozessor:** Welchen Prozessortyp besitzt der User?
- Cookies:** Hat der User auch tatsächlich die Cookies aktiviert oder besitzt er nur einen cookiefähigen Browser?
- 5
- Java:** Welche Version von Java kann auf dem Browser des Users ausgeführt werden?
- VB:** Ist die Performance des Users in der Lage VB auszuführen?
- 10
- Links:** Wie werden die Verlinkungen genutzt? Die Messung erfolgt bidirektional, d.h. es werden abgehende und eingehende Links gemessen und verglichen.
- 15 **Direkteinwahl:** Wie viele User wählen die Domain direkt an und wie ist die prozentuale Verteilung zu den anderen Einwahlmöglichkeiten?
- Sub-Domain:** Wie viele User kommen über eine evtl. untergeordnete Domain auf die Haupt-Domain? Diese Möglichkeit wird häufig genutzt, um den Erfolg einer Aktion zu bewerten.
- 20
- Neu-/Altkunde:** War der Besucher schon einmal auf der Domain, wie oft kommt er in welchen Zeitintervallen erneut vorbei? Welches Profil besitzt er und was sind seine Informationsbedürfnisse ?
- 25
- Verweilzeiten:** Welche Benutzergruppe verweilt wie lange bei welchen Informationen? Wie lange bleibt der Besucher auf bestimmten Seiten und wie lange dauert sein gesamter Besuch?
- 30 **Anzahl der Besuche:** Wie viele Besuche hatte die Domain und von wie vielen Besuchern wurden die Visits ausgelöst?
- Aktionen:** Die Aktionsauswertung ist ein wesentlicher Faktor des Verfahrens. Bei dieser Auswertung wer-

den alle Aktionen, die einen Verbesserungsprozeß auslösen sollen protokolliert und anschließend auf ihren Erfolg überprüft.

**Mail-Listings:** Protokollierung sämtlicher eingehender  
5 Mails

**Alarmsystem:** In diesem Modul werden die Eingriffsgrenzen definiert. Hier soll vermieden werden, daß Seiten verweisen und nicht die nötige Beachtung im Evaluierungsprozeß finden.

10 **Zielvorgaben-Kontrolle:** Dieses Modul dient der Zielvorgaben-Kontrolle. Hier werden die angestrebten Zielvorgaben mit dem erreichten Ziel verglichen und von den normal zu erwartenden Steigerungsraten bereinigt. Das Ergebnis ist eine Über- bzw. Unter-  
15 deckung der Zielvorgaben.

**e-commerce:** Dieses vergleichende Tool mißt das Verhältnis zwischen der Betrachtung von Informationen und den tatsächlichen Bestellungen. Dieses Verhältnis ist für weitere Optimierungsmaßnahmen im  
20 Kommunikationsmix ein entscheidender Planungsfaktor.

**Pikanalyse:** Die Pikanalyse stellt alle Vorgänge dar, die außerhalb eines Trends liegen. Durch die Untersuchung von ungewöhnlichen Ausschlägen können  
25 Rückschlüsse auf Werbemaßnahmen außerhalb des Webs getroffen werden.

Diese schon sehr ausführliche Dokumentation der historischen Daten, die natürlich von Interpretationsfehlern  
30 bereinigt sind, ist für eine zielorientierte Planung nicht ausreichend. Hier kommt die vergleichende Analyse zum tragen, die zwischen relativem und absolutem Erfolg unterscheidet. Natürlich können die zu vergleichenden Zeiträume frei gewählt werden. Ebenso werden evtl. Ak-

tualisierungen, die die Auswertung verfälschen könnten, bereinigt dargestellt.

Das Verfahren zeichnet sich deshalb durch folgende zusätzliche Funktionalitäten aus:

**Vergleich beliebiger Zeiträume:** Beim Vergleich beliebiger Zeiträume werden alle Übersichtsdaten in den entsprechenden Zeitintervallen dargestellt.  
10 Die daraus resultierenden Ergebnisse zeigen den Erfolg verschiedenster Maßnahmen auf. Die Ergebnisse sind von den Aktualisierungen und damit verbundenen möglichen Fehlinterpretationen bereinigt.

**Steigerungsraten Kompensation:** Jede Internet-Domain erfährt eine natürliche Steigerungsrate. Dies liegt häufig an der steigenden Anzahl von Altkunden, die immer wieder die Domain besuchen bzw. an der steigenden Zahl von Internet- Usern. Diese natürliche Steigerungsrate, die nicht unbedingt die  
15 Erfolgssteigerung widerspiegelt, wird im Verfahren ermittelt und kompensiert. So wird der exakte Erfolg ermittelt, der von möglichen Fehlinterpretationen bereinigt ist.  
20

**Analyse der Informationsbedürfnisse der User:** Die vergleichende Analyse erlaubt es, Änderungen der Informationsbedürfnisse der User festzustellen.  
25 Mit diesem Wissen kann der Content dem Userprofil angeglichen werden und so die Kundenzufriedenheit positiv beeinflusst werden. Dies ist insbesondere für das Binden von Altkunden von großer Bedeutung.  
30

**Analyse der relativen Ausgewogenheit:** Zum Start einer Internet-Domain wird man feststellen, daß User nicht die gesamte Bandbreite der Informationen benötigen. Durchschnittlich dauert der erste Besuch

einer Page 4 Minuten. Aus diesem Grund ist es wichtig Benutzerprofile anzulegen, um daraus eine Ausgewogenheit der gesamten Domain zu erzeugen. Die symmetrische Verteilung von Visits auf contentführenden Seiten ist anzustreben.

**Analyse der Veränderungen von Rechnerfähigkeiten:**

Möchten Sie auch gerne wissen, ob Ihr Klientel die Cookies aktiviert hat oder nicht? Oder ob Sie endlich einmal eine andere Auflösung als Standard voraussetzen können? Mit der vergleichenden Analyse sehen Sie auf einen Blick die Veränderungen der Performance von Rechnern. Dies bietet genügend Raum rechtzeitig auf die veränderten Bedingungen reagieren zu können.

**Suchmaschinen:** Das Eintragen in Suchmaschinen ist sehr wichtig. Oft ist dies die einzige Möglichkeit Informationen zu beschaffen. Es reicht nicht aus, Anzeigen mit www. Adresse zu schalten oder diese auf der Visitenkarte zu erwähnen. Ein Eintrag auf den vorderen Plätzen ist von immenser Bedeutung. Dazu müssen die Erfolge der Suchmaschinen und die Suchworte analysiert und die Veränderungen abgespeichert werden.

**Verlinkungen:** Wie viele neue Besucher Ihrer Seite resultieren aus einer Verlinkung? Ist der Werbeträger geeignet für eine Plazierung von Werbebotschaften? Diese und weitere Fragen werden mit diesem Modul beantwortet.

**Alt-/Neukunde:** Sehr interessant sind die Veränderungen des Neu-/Altkunden Verhältnisses. Ist es möglich Kunden dauerhaft durch einen besonderen Service an die Domain zu binden? Ein ausgewogenes Verhältnis zwischen Neu-/Alt-kunden ist von großer Wichtig-

keit. Es zeigt auf, wieviel Vertrauen Sie zu Ihrem Kunden via Internet aufbauen können.

**Verweilzeiten:** Wie verändern sich die Verweilzeiten besonders im Altkunden-Segment? Die Antwort auf diese Frage drückt die Entspanntheit und das Vertrauen in den Informationsgehalt aus. Ebenso sind die Veränderungen der Userpfade wichtig für eine kundennahe Gestaltung der Domain.

**Steigerungsraten im e-commerce:** Die Veränderungen zwischen Informationsbeschaffung und der tatsächlichen Bestellung von Dienstleistungen, Informationen oder Produkten sind wichtige Kennzahlen für die Absatzplanung. Die hieraus resultierenden Ergebnisse lassen sich direkt in den entsprechenden Planungswerkzeugen einsetzen.

**Vergleichende Pikanalyse:** Abweichungen im normalen Zugriffsverhalten, die z.B. durch Anzeigen oder andere Aktionen ausgelöst werden, können hier genau verfolgt werden. Als Beispiel dient hier eine Reihe von Anzeigen, die im 14tägigen Abstand unverändert in derselben Zeitung erscheint. Nach einer gewissen Laufzeit werden die Ausschläge im Internet immer geringer. Dies kann als eindeutiges Zeichen für das „Abnutzen“ des Inserates gewertet werden. Die Konsequenz: Die Anzeigen-Aktion pausieren lassen oder eine neue Anzeige konzipieren.

Nachdem das Analyseverfahren richtige und bereinigte Daten ermittelt hat und diese in eine vergleichende Analyse münden, ist die logische Konsequenz die Ermittlung und Berechnung der Trends. Diese Trendberechnung wird nach DIN ISO durchgeführt. Als Basis werden statistische Modelle der Qualitätssicherung angewendet. Die-

se Berechnungen sind im Einklang mit der DGQ und dem TÜV cert entstanden.

Beantwortet werden folgende Fragen :

- 5 - Wie entwickeln sich die Seitenzugriffe?
- Wie ändert sich zukünftig das zeitbezogene Userverhalten?
- Wie entwickelt sich die Rechner Performance?
- Welchen Einfluß werden zukünftig Suchmaschinen und  
10 Links auf den Erfolg einer Domain haben?
- Wie wird sich zukünftig das Verhältnis Neu-/Altkunde verändern und welche Auswirkungen wird dies haben?
- Wie wird sich zukünftig das Verhältnis zwischen  
15 Informationsbedarf und Bestellungen im e-commerce Sektor ändern?

Ferner benötigt der Webmaster einen Referenzpunkt, um seine eigene Position bestimmen zu können. Es stellt  
20 sich z.B. die Frage: "Ist das Verhältnis von 36:1 von Informationsbeschaffung zu tatsächlicher Bestellung positiv oder negativ zu bewerten ? Diese Kalibrierung erfolgt durch eine externe Erfahrungsdatenbank.

In der externen Erfahrungsdatenbank wird das Know-how  
25 von vielen Domains komprimiert dargestellt. Sämtliche Daten sind von Erkennungsmerkmalen der Firmen bereinigt. Diese Daten werden den Usern des Verfahrens online zur Verfügung gestellt. Hier werden Informationen über Trends im Internet, Erfahrungen im Bereich e-commerce,  
30 typische User-Profile sowie Informationen über das Userverhalten gespeichert sein. Aus diesen Erfahrungen resultieren diverse Optimierungsmöglichkeiten für die vorhandene Internet Domain.



Zur Optimierung der Domain stehen diverse Tools zur Verfügung, die sich in drei generelle Kategorien einteilen lassen :

Kategorie 1 beinhaltet statische Module. Diese Module optimieren den Code der Domain in zweierlei Hinsicht. Zum einen werden die Module zur exakten Analyse integriert. Zum anderen werden mögliche Fehler kompensiert und korrigiert. Es ist ein „Übersetzer“ integriert, der die technische Information in eine anschauliche, dem Content entsprechende Darstellungsform transferiert. In diesem Abschnitt werden die Parameter und Zielvorgaben gesetzt, Fragen nach Eingriffsgrenzen gestellt und so das Verfahren initialisiert. Die in der Kategorie 1 befindlichen statischen Module sind:

- Codeoptimierung
- Frame-Korrektur
- Proxy-Optimierung
- Struktur-Konverter
- allgemeine Korrekturmodule
- Parameter Modul

Kategorie 2 beinhaltet dynamische Module, die Informationen während des Betriebes der Internet-Domain liefern. Zu diesen Modulen zählen die implementierten Statistik-Module und ein sogenannter Error-Messenger. Dieser schickt bei Auftritt eines Fehlers auf dem Server ein E-Mail an den Webmaster mit dem Vermerk, welcher Fehler beim User entstanden ist.

Der User erhält im Gegenzug die Meldung, daß dieser Fehler schnellstmöglich beseitigt wird. Diese Maßnahme ist ein entscheidender Schritt in Richtung Kundenorientierung.

Kategorie 3 beinhaltet dynamisch-administrative Module. Der Suchwort-Optimierer analysiert aus den Contentsei-

- ten die Wörter, die eine hohe Frequenz aufweisen und erarbeitet daraus eine Vorschlagsliste. Diese Vorschlagsliste wird vom Webmaster angenommen bzw. verbessert. Diese Wörter werden anschließend in das Modul Meta-Optimierung einprogrammiert. Das Resultat ist ein Metatag, der allen Anforderungen der Suchmaschinen entspricht und den größtmöglichen Erfolg garantiert. Die Veränderungen werden im Suchmaschinen-Eintragsmodul gespeichert. Dieses Modul nimmt automatisch den Kontakt zu den gewählten Suchmaschinen auf und veranlaßt die Suchrobots zu einem Besuch der aktualisierten Domain. Durch diese Maßnahme ist es erstmals möglich diese Aufgabe teilweise zu automatisieren. So ist es möglich, das Ranking wesentlich zu verbessern und eine der oberen Positionen permanent beizubehalten. Weitere Module generieren Vorschlagslisten für sämtliche Aktionen, die offline betrieben werden müssen. Zu diesen zählt z.B. ein Vorschlagswesen für Anzeigenschaltungen.
- Der nun innenliegende Regelkreis ist geschlossen: Messung, Vergleich, Trendberechnung, Optimierung, Übertragung und erneute Erfolgsmessung. Dieses Verfahren ermöglicht es Schritt für Schritt einen kontinuierlichen Verbesserungsprozeß einzuleiten, welcher von der DIN ISO 9000 gefordert wird.

Ein weitere Bestandteil des Verfahrens ist ein Lernprogramm für effektives Webcontrolling und ein Configurations-Tool. Dieses Tool vermittelt multimedial sämtliches Know-how, welches für den sukzessiven Aufbau und die Vermarktung von Content im Internet relevant ist. Hier können folgende Informationen abgerufen werden:

Was kann analysiert werden?

Wie sicher sind die Aussagen und die Daten?

Welche Datenschutzgesetze müssen beachtet werden?  
Welche Punkte müssen besondere Beachtung finden?  
Wie kann ein kontinuierlicher Verbesserungsprozeß  
erreicht werden?  
5 Wie wird ein effektives Webcontrolling am bestens  
strukturiert?

Im Modul „Configuration Tool“ werden dem Webmaster konkrete Fragen zu seinem Informationsbedürfnis gestellt.  
10 In diesem Bereich entsteht die Darstellungsform der Ergebnisse. Es erfolgt eine Nachbildung der Contentstruktur sowie eine Vorstellung des Verfahrens als Basis für das effektive Webcontrolling. Dieses Tool modifiziert und optimiert den Quellcode einer Domain. Weiterhin werden strukturelle Änderungen durchgeführt und  
15 bestimmte Merkmale hinzugefügt. Diese Modifikationen werden gesichert, so daß zu jedem Zeitpunkt der Originalzustand wieder hergestellt werden kann. Aus diesem Grund ist der zentrale Datenpool das Herzstück des Verfahrens.  
20

In einem weiteren Modul, dem zentralen Datenpool, wird eine Sicherung der Domain vor jeglicher Veränderung angelegt. Dort wird neben der alten Struktur der Daten  
25 die nun modifizierte Struktur angelegt. Die Parameter und die historischen Änderungen werden abgespeichert. So ist es möglich, bei der vergleichenden Analyse Weiterentwicklungen zu integrieren. Fehlinterpretationen sind so nahezu ausgeschlossen. Parallel zum klassischen  
30 Logfile erfolgt die Etablierung eines zweiten Logfiles. Dieser Logfile wird neben dem externen Logfile gespeichert. Sämtliche Daten sind nun unabhängig vom Provider zu führen; automatisches Downloaden von Logfiles ermöglicht das lückenlose Erfassen von Userdaten. Hier wer-

den sämtliche Analyse-Auswertungen gespeichert und in verständlicher Form im Analysebereich webmastergerecht dargestellt. Diese Informationen können auch über die Universalschnittstelle in bereits existierende Planungswerkzeuge integriert werden. Natürlich können auch diese Daten in individuellen Erweiterungsmodulen ihren Einsatz finden.

Die offene Plattform, die das Verfahren darstellt, erlaubt es weitere Werkzeuge zu kreieren. Diese Module werden individuell nach Kundenwunsch gefertigt. Zu diesen möglichen Produkten zählt das Mail-Controlling, das Kommunikationsmix-Controlling oder das Administrations-Controlling.

Der rasante Fortschritt in den modernen Kommunikationstechniken, wie z.B. ADSL, HDSL und power line communication ermöglicht ungeahnte multimediale Fähigkeiten im Internet. Die Inhalte werden stärker auf zielgruppenspezifische Aspekte ausgerichtet. Interaktivität fördert den Dialog mit den einzelnen Kundensegmenten. Somit können Service- und Dienstleistungsmerkmale besser hervorgehoben werden. Internet ermöglicht eine kostengünstige, kontinuierliche Kommunikationskette zum Kunden und bildet eine breitgefächerte Informationsbasis ohne unmittelbaren Kaufdruck. Die Berechnung des Erfolges bzw. return on investment (ROI) einer Internetseite ist entscheidend für die strategische und inhaltliche Weiterentwicklung des Content, der Kundenansprache und der Kundennähe.

Ziel des Verfahrens ist - im Sinne eines hochwertigen Reporting Systems - die Erhöhung der Interaktivität zum Kunden. Die Herausstellung von Produkt- bzw. Service-

vorteilen sowie die Optimierung der Logistik und des Kostenmanagements sind ebenfalls von hoher Relevanz.

5 Einfache Web-Site-Statistiken lassen sich durch Standard-Analysen erstellen. Diese bleiben jedoch generell unspezifisch, da sie keine Identifikation des Users und der User-Pfade sowie keine Analyse der Vertiefungsgrade der Abfrage, der Pausen, der Verweildauer etc. ermöglichen. Attraktive Inhalte und Kommunikationsinstrumente  
10 lassen sich jedoch nur durch eine personifizierte Analyse der Benutzerbedürfnisse realisieren. Daher ist die Zielgruppenanalyse, also die Unterteilung nach typischen Verhaltensmerkmalen, von immenser Bedeutung.

15 Um neue, innovative Dialog-Konzepte zu generieren, reicht die Standard-Analyse nicht aus. Nur eine Kombination verschiedener Methoden ermöglicht ein idealtypisches Dialog-Konzept und gewährleistet, daß alle wesentlichen Einflußfaktoren einer zukunftsweisenden und  
20 qualifizierten Kundenorientierung Berücksichtigung finden.

Der so zu realisierende Nutzen für den Anwender läßt sich wie folgt spezifizieren:

- 25
- Flexibilisierung und Optimierung des kundenorientierten Kommunikationsmix
  - Dokumentation des Internet-Auftritts
  - Erfolgsmessung und Offenlegung von Stärken und potentiellen Herausforderungen der Kommunikation

30

  - Etablierung integrierter Kommunikationskonzepte durch Print-Medien, TV, Hörfunk,
  - Direct Mailing und elektronische Medien

- unmittelbare Ableitung von Markt- und Nachfrage-trends und sofortige Anpassung des Kommunikations-mix
- Senkung der Medienbelegungs-, Logistik- und Print-kosten
- optimierte Zielgruppenansprache.

Der Funktionalitätsumfang des Verfahrens eröffnet ihm Einsatzmöglichkeiten, die vom privater Anbieter über lokale und regionale Anbieter bis hin zu national/multinationalen Anbietern reichen.

Der private Anbieter stellt auf einem virtuellen Webspaces persönliche Informationen zur Verfügung. Wie der private Anbieter verfügt der lokale Anbieter häufig über keinen eigenen Server. Sein Anforderungsprofil entspricht somit weitestgehend dem des privaten Anbieters. Beide werden mit nur einigen der beschriebenen Module auskommen, um ihre Anforderungen zu realisieren. Der regionale Anbieter stellt ein weitaus umfangreicheres Informationsangebot zur Verfügung, da er zumeist über mehrere Dependancen in einer Region verfügt. Es ist davon auszugehen, daß dieser Kunde über ausgereifere technische Möglichkeiten verfügt, um verifizierte Daten über das User-Verhalten zu erhalten. Er ist somit potentieller Anwender des Verfahrens im vollen Umfang. Die national/multinationalen Unternehmen verfügen zumeist über einen eigenen Server und sind somit in der Lage, das gesamte Internet-Spektrum auszunutzen. Sie sind potentielle Kunden des gesamten Verfahrens einschließlich der Trainings-Tools. Je nach Unternehmensgröße empfiehlt sich außerdem die individuelle Programmierung von Schnittstellen zu anderen handelsüblichen Warenwirtschaftsprogrammen und Planungswerkzeugen.

Das Dienstleistungskonzept beginnt mit der Ist-Analyse des Internetauftrittes des Kunden. Hierbei wird eine Stärken-/Schwächen-Analyse vorgenommen, die kundenseitige Verbesserungspotentiale aufzeigt.

5

Da nicht alle Features des Controlling Tools automatisiert werden können, werden einige Tools manuell entsprechend dem Anforderungsprofil des Kunden adaptiert.

10

Hieraus resultieren für den Kunden diverse Vorteile. So wird durch das "Outsourcen" von Software-Spezialisten jederzeit ein qualitativ hochwertiger Support sichergestellt und gleichzeitig eine Kostenreduktion im Bereich des Personalwesens des Kunden realisiert. So ist Spezialisten Know-how "on demand" abrufbar oder alternativ eine vor-Ort-Betreuung durch einen lokalen Spezialisten - z.B. einen Franchisenehmer- möglich.

15

20 Ein optimales Dienstleistungskonzept beinhaltet folgende Komponenten:

- Ist-Analyse
- Workshop zur Determinierung von Verbesserungspotentialen
- 25 - Festlegung des Soll-Zustandes
- Zieldefinition : kurzfristig - langfristig
- Umsetzung mit dem erfindungsgemäßen Verfahren
- Kontrolle

30 Per Ist-Analyse wird der aktuelle Zustand der Domain determiniert und alle verwertbaren Informationen aus dem Logfile entnommen. Anschließend erfolgt eine Potentialermittlung in einem Workshop in Kooperation mit dem

Kunden. Ferner werden hier der Soll-Zustand sowie die kurz- und langfristigen Ziele festgelegt.

Durch die Implementierung der einzelnen Module des Verfahrens, der Meta-Optimierung und der strukturellen Optimierung werden erste fehlerbereinigte Analysen des User-Verhaltens ermöglicht. Anhand historischer Daten erfolgt die Durchführung von Trendberechnungen, die als Basis für die weitere Planung dienen. Anschließend erfolgt eine kundengerechte Aufbereitung der Daten und daraus resultierend Optimierungsvorschläge.

Diese Ergebnisse werden mit den Daten der externen Datenbank sowie mit den Daten des durch das Verfahren angelegten Datenpools abgeglichen und evaluiert. Nach Präsentation dieser Informationen erfolgt ein erneuter Abgleich der kurz- und langfristigen Ziele und der daraus resultierenden Maßnahmen in Zusammenarbeit mit dem Kunden.

In Übereinstimmung mit diesen Zielen erfolgt nun die Content- und Strukturoptimierung sowie die Entwicklung der Corporate Identity und die Optimierung des Workflows. Hiermit schließt sich der Regelkreis, der in beliebigen Intervallen von 1 - 3 Monaten wiederholt wird.

25

Das Verfahren umfaßt folgende spezielle Dienstleistungen:

- Optimierung des Kommunikationsmixes durch gezielte Werbewirksamkeitsanalysen
- 30 - Etablierung von Sponsor-Konzepten durch exakte Analyse von Besucherzahlen und Zielgruppen
- Online-Werbung incl. Erfolgsmessung und Kosten-/Nutzen-Analyse



- Unterstützung von "Star-Allianzen" durch Schnittstellen-Programmierung zu Warenwirtschaftssystemen und Etablierung von Logistik-Konzepten
- Optimierung des Workflows von Bestellung bis zur  
5 Lieferung und Fakturierung in den jeweiligen Partnerunternehmen

Folgende Charakteristika unterstreichen die Einzigartigkeit des Dienstleistungs-konzepts für den Kunden :

- 10 - durch Outsourcing ressourcenschonende Personal- und Kostenstruktur
- innovative, "state-of-the-art"-Lösungen durch permanent geschultes Fachpersonal
- schnelle, messbare Erfolge durch zeitnahes Etablieren von Konzepten  
15
- unabhängige und wertfreie Analyse des Userverhaltens durch qualitätsgesicherte Maßnahmen
- eindeutige Präsentation der Ergebnisse für die Entscheidungsträger
- 20 - unternehmensunabhängige Auswertung der Analyseergebnisse als "Stabsfunktion"
- Investitionsminimierung und -schutz durch sukzessive Investitionstätigkeit mit paralleler Erfolgsmessung
- 25 - Netzwerke durch Kooperation mit Marktführern und Know-how-Trägern zur Lösung individueller Herausforderungen

Die Erfindung ist nicht beschränkt auf die hier dargestellten Ausführungsbeispiele. Vielmehr ist es möglich,  
30 durch Kombination und Modifikation der genannten Mittel und Merkmale weitere Ausführungsvarianten zu realisieren, ohne den Rahmen der Erfindung zu verlassen.

### Patentansprüche

1. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz,  
5 dadurch gekennzeichnet, daß
  - nach Implementierung des Verfahrens der Quellcode der Domain optimiert wird,
  - von mindestens einem Servercomputer Benutzerakti-  
10 vitäten erfaßt und protokolliert werden,
  - die erfaßten Aktivitäten von Interpretationsfehlern bereinigt werden,
  - 15 - die nach der Bereinigung vorliegenden Daten analysiert und mit Daten aus einer Erfahrungsdatenbank verglichen werden und
  - 20 - die Ergebnisse des Vergleiches zur Optimierung eingesetzt werden, indem durch eine Rückführung eine aktive Einflußnahme auf den Servercomputer erfolgt und hierdurch ein geschlossener Regelkreis realisiert wird.
- 25 2. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach Anspruch 1,  
dadurch gekennzeichnet, daß  
30 die Benutzeraktivitäten in einem Logfile protokolliert und/oder in einer zweiten Datenbank gespeichert werden.

3. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach Anspruch 1 oder 2,  
dadurch gekennzeichnet, daß
- 5 im Logfile und/oder der zweiten Datenbank
- Parameter des Controllings und/oder
  - Daten über die Struktur der HTML-Dokumente und/oder
  - die Ist-Analyse und/oder
  - 10 - die historischen Daten und/oder
  - die Trend-Analyse
- gespeichert werden.
4. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 3,  
dadurch gekennzeichnet, daß
- 15 das Verfahren eine Universalschnittstelle zur Verfügung stellt.
- 20
5. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 4,  
dadurch gekennzeichnet, daß
- 25 die protokollierten Benutzeraktivitäten über einen Zeitraum von mehreren Monaten gespeichert werden.
- 30 6. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach Anspruch 1,

dadurch gekennzeichnet, daß

die Erfahrungsdatenbank Informationen enthält

- über alle HTML-Dokumente, die, versehen mit Attributen, in Zuordnungs-Tabellen aufgelistet sind, und/oder
- über Parameter des Controlling.

7. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 6,

dadurch gekennzeichnet, daß

die den HTML-Dokumenten zugeordneten Attribute die HTML-Dokumente als

- technische Dokumente,
  - Dokumente zur Navigation,
  - contentführende Dokumente oder
  - werbeführende Dokumente
- kennzeichnen.

8. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 7,

dadurch gekennzeichnet, daß

die durch den Aufbau der Website und/oder

die durch den Einsatz von

- Proxy-Cacheservern und/oder
- Lokalen Cachespeichern und/oder
- Routern und/oder
- Firewalls und/oder
- dynamisch vergebenen IP-Adressen

hervorgerufenen Interpretationsfehler bereinigt werden.

9. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß
- 5 jede Anfrage an eine dynamisch erzeugte Seite einschließlich der generierten Inhalte protokolliert wird.
10. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen nd zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß
- 10 jede Anfrage an eine Website zu einem Eintrag in das Logfile führt, indem eine Anfrage mittels eines CGI-Images immer dem Server mitgeteilt wird.
- 15 11. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß
- 20 die Bereinigung der Interpretationsfehler mit Hilfe von Zuordnungstabellen erfolgt.
12. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 11, dadurch gekennzeichnet, daß
- 25 - die automatische Behebung von Programmierfehlern in den HTML-Dokumenten und/oder

- die automatische Anpassung der HTML-Dokumente an die gerätetechnischen Voraussetzungen der Nutzer und/oder
  - die automatische Anpassung der Inhalte der Websites an das Nutzerverhalten und/oder die Anforderungen der Nutzer
- Bestandteil der Optimierung sind.

13. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 12,
- dadurch gekennzeichnet, daß
- die Optimierung durch
- Vergleich mit dem typischen Zielgruppenverhalten und/oder
  - nach Methoden des kundenbezogenen Qualitätsmanagements
- erfolgt.

14. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 13,
- dadurch gekennzeichnet, daß
- eine Darstellung der Ergebnisse der Analyse nach Vorgabe individueller Parameter erfolgt.

15. Methode zur Optimierung der Web-Präsenz in Computernetzen,
- dadurch gekennzeichnet, daß
- das Benutzerverhalten erfasst und analysiert wird, wobei fehlerhafte Einflußfaktoren eliminiert werden

und eine Optimierung der Web-Präsenz mit dem Ziel der Erhöhung der betriebswirtschaftlichen Wirksamkeit der Web-Präsenz erfolgt derart, daß aus den erfassten, bereinigten Benutzeraktivitäten und Aussagen aus einer Erfahrungsdatenbank ein Regelwert zur Realisierung eines geschlossenen Regelkreises gebildet wird, welcher Einfluß auf die Gestaltung der Web-Präsenz ausübt.

**THIS PAGE BLANK (USPTO)**



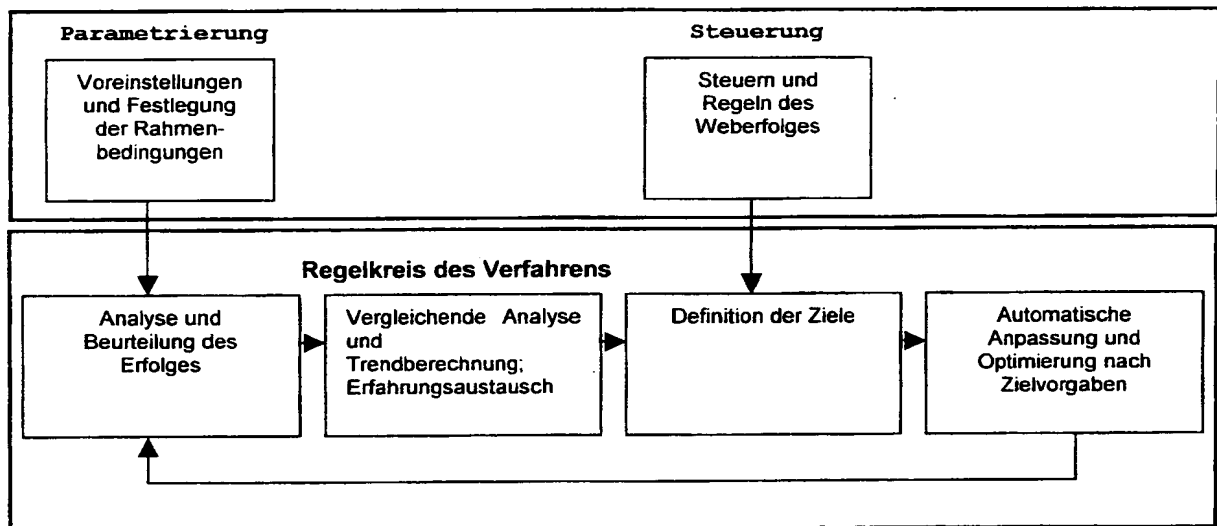


Fig. 1

**THIS PAGE BLANK (USPTO)**

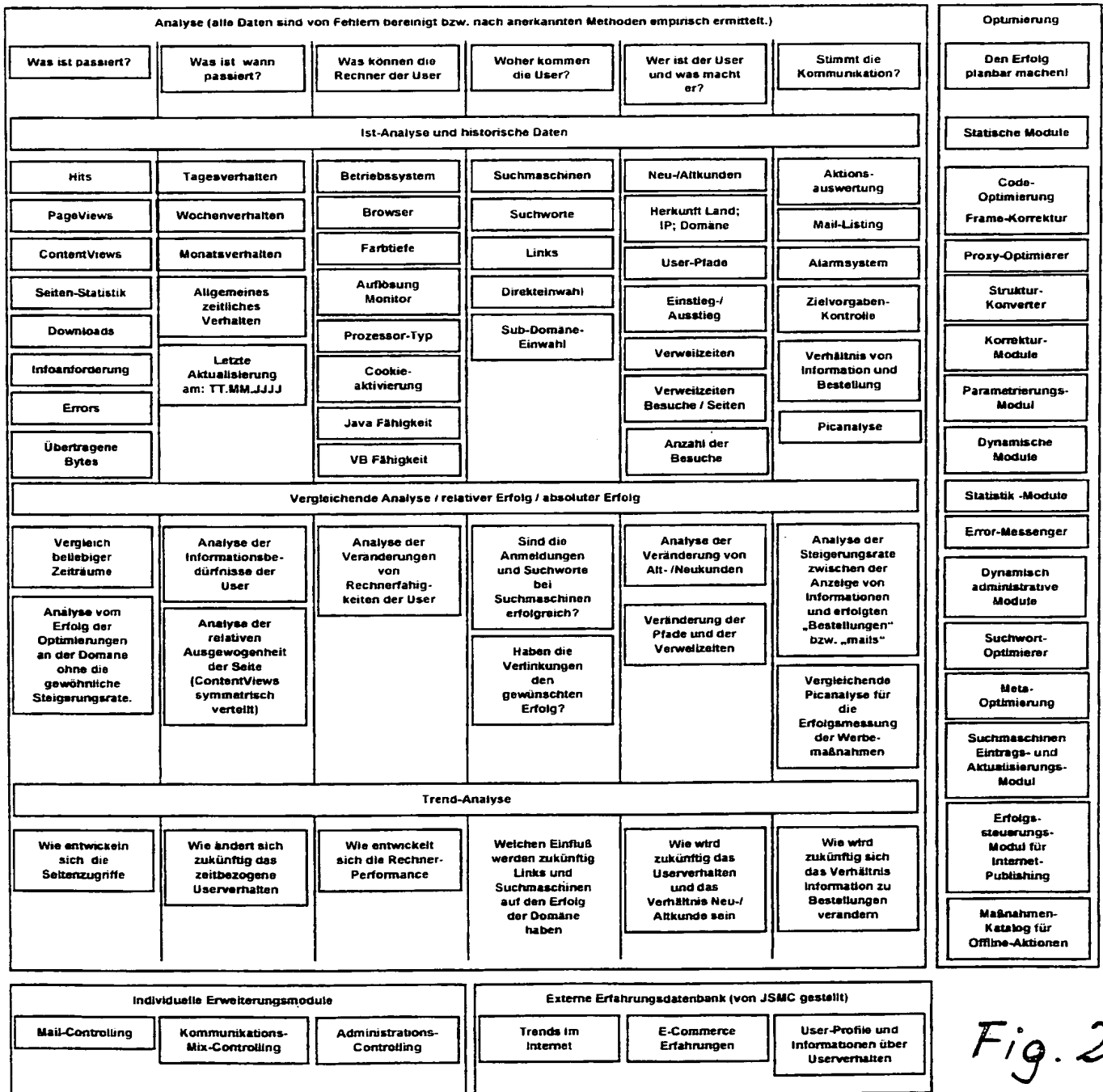
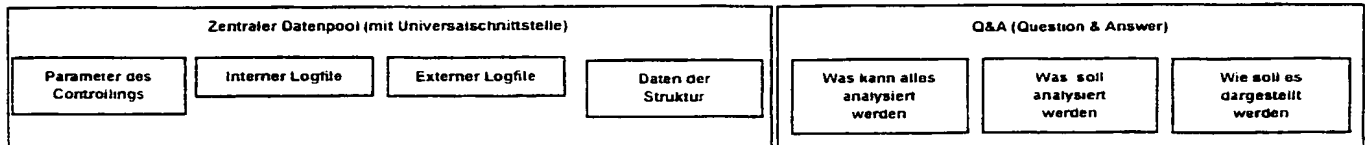
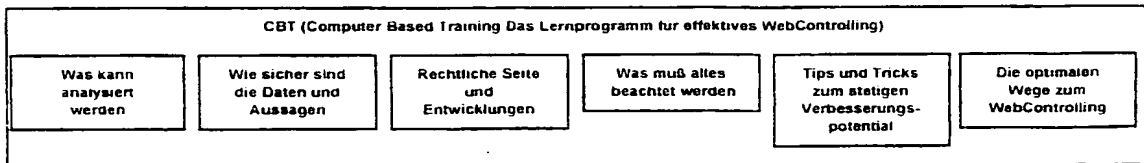
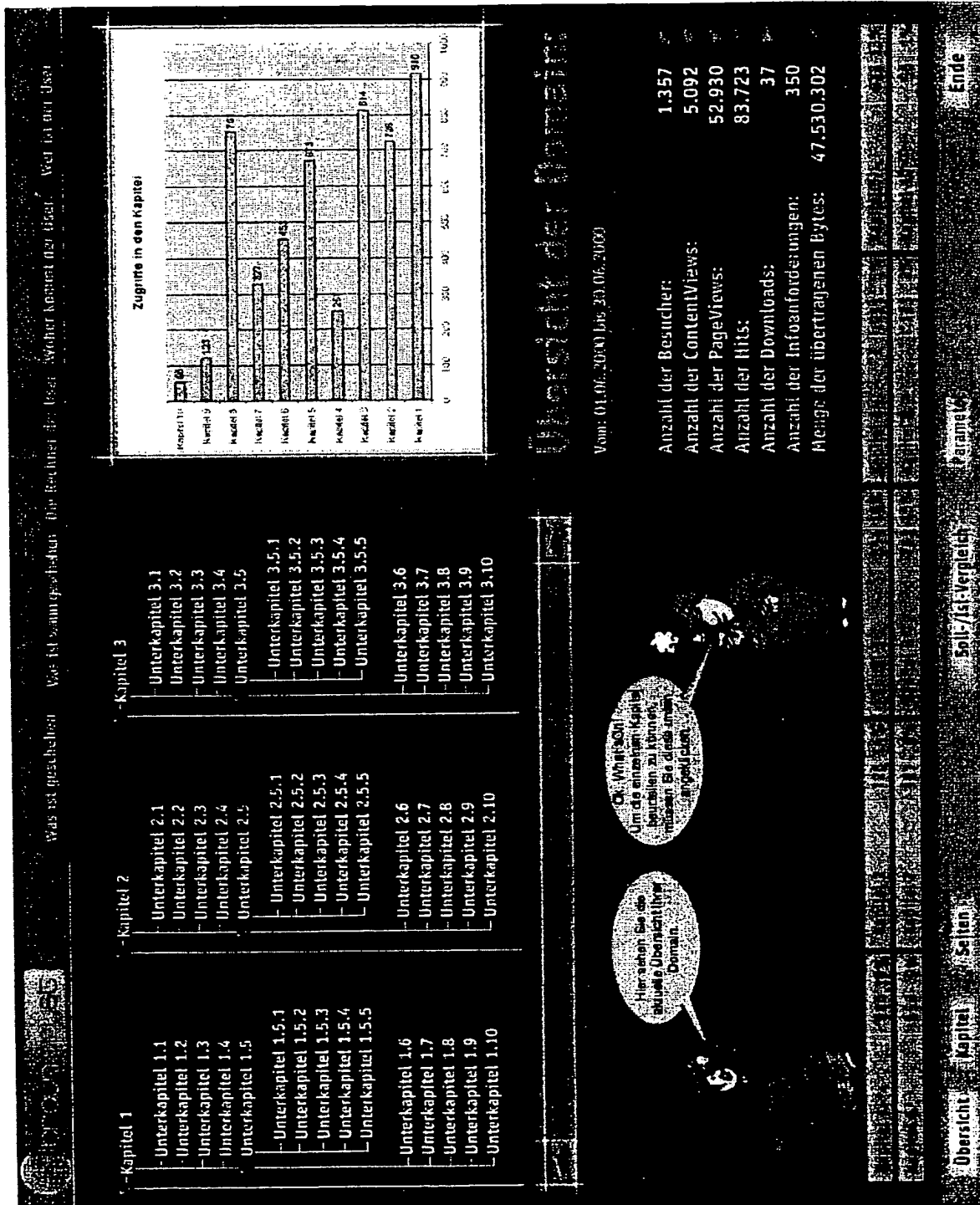


Fig. 2

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**

09/868110

m f

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

## BERICHTIGTE FASSUNG

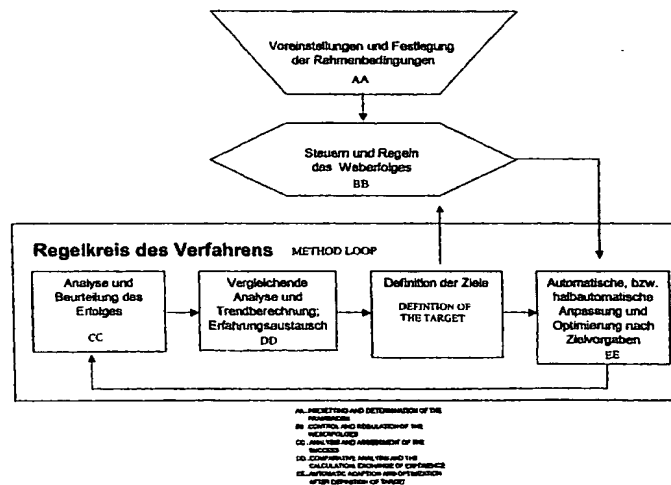
(19) Weltorganisation für geistiges Eigentum  
Internationales Büro(43) Internationales Veröffentlichungsdatum  
19. April 2001 (19.04.2001)

PCT

(10) Internationale Veröffentlichungsnummer  
WO 01/027760 A3

- (51) Internationale Patentklassifikation<sup>7</sup>: G06F 17/60 (71) Anmelder und  
(72) Erfinder: SCHLAFKE, Joachim [DE/DE]; Hauptstr. 63,  
53859 Niederkassel (DE).
- (21) Internationales Aktenzeichen: PCT/EP00/10136
- (22) Internationales Anmeldedatum: 13. Oktober 2000 (13.10.2000) (74) Anwälte: HENGELHAUPT, Jürgen, D. usw.; Gulde  
Hengelhaupt Ziebig, Schützenstrasse 15 - 17, 10117 Berlin  
(DE).
- (25) Einreichungssprache: Deutsch (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,  
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität: 199 51 209.4 15. Oktober 1999 (15.10.1999) DE

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR THE ANALYSIS OF USER BEHAVIOR IN A COMPUTER NETWORK IN ORDER TO OPTIMIZE  
WEB PRESENCE(54) Bezeichnung: VERFAHREN ZUR ANALYSE DES BENUTZERVERHALTENS IN COMPUTERNETZEN ZUR OPTIMIE-  
RUNG DER WEB-PRÄSENZ

(57) Abstract: The invention relates to a method for the analysis of user behavior in computer networks in order to optimize web presence. Said method can in particular be used to obtain qualified and future-orientated data on user behavior on the internet and intranet in order to influence one's own internet presence as a means for planning production and turnover. Said method provides a feedback circuit comprising an actual status analysis, designation of target functions, constant observation of user behavior, actual/desired comparison, performance of a trend calculation in addition to content and structure optimization of said web presence.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz, welches insbesondere eingesetzt werden kann, um qualifizierte und zukunftsorientierte Daten über das Userverhalten im Internet und Intranet zu erhalten, um damit

[Fortsetzung auf der nächsten Seite]

WO 01/027760 A3



MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(48) Datum der Veröffentlichung dieser berichtigten

Fassung:

6. September 2002

(84) Bestimmungsstaaten (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(15) Informationen zu Berichtigungen:

siehe PCT Gazette Nr. 36/2002 vom 6. September 2002, Section II

Frühere Berichtigung:

siehe PCT Gazette Nr. 31/2001 vom 2. August 2001, Section II

Veröffentlicht:

— mit internationalem Recherchenbericht

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(88) Veröffentlichungsdatum des internationalen Recherchenberichts:

11. Juli 2002

Einfluss nehmen zu können auf die eigene Internet-Präsenz als Mittel zur Produktions- und Umsatzplanung. Das Verfahren stellt einen Regelkreis zur Verfügung, welcher besteht aus einer Ist-Analyse, der Vorgabe von Zielfunktionen, der ständigen Beobachtung des Userverhaltens, dem Ist/Soll-Vergleich, einer Durchführung der Trendberechnung und der Content- und Strukturoptimierung der Web-Präsenz.



---

**Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz**

---

5

Die Erfindung betrifft ein Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz, welches insbesondere eingesetzt werden  
10 kann, um qualifizierte und zukunftsorientierte Daten über das Userverhalten im Internet und Intranet zu erhalten, um damit Einfluß nehmen zu können auf die eigene Internet-Präsenz als Mittel zur Produktions- und Umsatzplanung.

15

Seit der Einführung des Internet (world wide web oder auch www) ist die Zahl der online-User weltweit enorm angestiegen. Durch diese Entwicklung ist das Internet zu einer wichtigen Plattform für die Wirtschaft auf internationaler Ebene geworden. Mit der fortschreitenden  
20 Kommerzialisierung wird das Netz immer multimedialer; Grafiken, Videos, Musik und Sprache ergänzen die reinen Textinformationen. Die ursprünglich auf reinen Informationsaustausch ausgerichtete Kommunikationskultur gestattet heute die Verbreitung von Werbebotschaften. Andere Medien wie Mobiltelefone oder TV-Geräte werden unmittelbar in den durch das Internet bereitgestellten Kommunikationsprozeß integriert. Somit ist Internet zu  
25 einem vollwertigen Kommunikationsinstrument im business-to-business (B2B) und ebenfalls im business-to-customer (B2C) Markt gereift. Dies führt zu einer drastischen Veränderung der Konsumgewohnheiten und des Konsumverhaltens der privaten Haushalte. Um den veränderten Anforderungen gerecht zu werden, ist es unter-

30

nehmensseitig unerlässlich, das Angebot an die veränderten Bedürfnisstrukturen der Konsumenten zu adaptieren. Innovative Unternehmen im Konsumgüter- und Dienstleistungsbereich erkennen das enorme Marktpotential und investieren in diese integrativen Kommunikationsplattformen. Um proaktiv die Zukunft dieses Mediums mitzugestalten, bedarf es detaillierter Informationen über die jeweiligen sich permanent modifizierenden User-Profile. Die Chancen durch ein Controlling-Instrument exakte User-Profile zu erstellen ist größer denn je, birgt aber auch die Gefahr der Fehlinterpretation der erhaltenen Daten.

Bei allen bekannten diesbezüglichen Verfahren wird nur quantitativ das Userverhalten protokolliert, basierend auf dem sog. Logfile, einem kontinuierlichen Logbuch, in welchen sämtliche Anfragen an den Server protokolliert werden. Ebenfalls, und das wird meist unterschlagen, wird eine Genauigkeit von den Systemen aufgezeigt, die bewiesenermaßen falsch ist. Das im Internet verwendet Verfahren des Proxy-Cash-Servers, welches eingesetzt wird um die Gesamtperformance des Internets zu erhöhen, führt beispielsweise zu keinem Eintrag im Logfile wenn die Seite über den Cash-Server geladen wird. In diesem Falle werden tatsächliche Besucher der entsprechenden Seite nicht registriert. Umgekehrt führt der Besuch von „ROBOTS“, also von Spionen der Suchmaschinen, zu Einträgen, die als Besucher der Website im Logfile protokolliert werden, aber im Sinne der Analyse keine sind. Sobald diese Situationen nicht berücksichtigt werden, ist die Interpretation mit Fehlern behaftet. Die Erklärung für diesen Zustand ist recht einfach. Der Logfile ist ein Format für die zeitliche Dokumentation des Webserverns. Er ist nicht für die Analy-

se des Userverhaltens direkt konzipiert. Aus diesem Grund ist das Ergebnis des Logfile nicht falsch sondern wird vielmehr inkorrekt interpretiert.

5 Der Erfindung liegt deshalb die Aufgabe zugrunde, ein Verfahren zu entwickeln, mit welchem die Mängel der bekannten Verfahren vermieden werden, welches insbesondere

- von Fehlern bereinigte Daten zur Verfügung stellt,
- 10 - eine separierte Darstellung von Daten, die eine Fehlertoleranz aufweisen, inkl. der Darstellung der Bandbreite der Toleranz,
- eine Unterscheidung in Neu- und Altkunden sowie
- die Clusterung in Kundensegmente erlaubt,
- 15 - zukunftsorientierte Daten und Planungsdaten liefert, welche durch
- beliebig wählbare Analyse-Intervalle ermittelt werden können,

und darüber hinaus dem User eine erweiterte technische  
20 Performance wie z.B. Sprache, Bildschirmauflösung, Farbtiefe etc. zur Verfügung stellt und durch die historische Dokumentation des Internetauftritts dem User ein Kontrollinstrument für den Kommunikationsablauf in die Hand gibt, welches ihn in die Lage versetzt,  
25 schnell das Nutzerverhalten im Internet zu analysieren und darauf basierend erfolgreich Planungsmaßnahmen durchzuführen.

Diese Aufgabe wird erfindungsgemäß gelöst durch die  
30 Merkmale im kennzeichnenden Teil des Anspruchs 1 im Zusammenwirken mit den Merkmalen im Oberbegriff. Zweckmäßige Ausgestaltungen der Erfindung sind in den Unteransprüchen enthalten.

Ein besonderer Vorteil der Erfindung liegt darin, daß durch das Verfahren das Benutzerverhalten in Computernetzen analysiert und darauf basierend die eigene Web-Präsenz optimiert wird, indem nach Implementierung des Verfahrens der Quellcode der Domain optimiert wird, von  
5 mindestens einem Servercomputer Benutzeraktivitäten erfaßt und protokolliert werden, die erfaßten Aktivitäten von Interpretationsfehlern bereinigt werden, die nach der Bereinigung vorliegenden Daten analysiert und mit  
10 Daten aus einer Erfahrungsdatenbank vergleichen werden und die Ergebnisse des Vergleiches zur Optimierung eingesetzt werden, indem durch eine Rückführung eine aktive Einflußnahme auf den Servercomputer erfolgt und hier-durch ein geschlossener Regelkreis realisiert  
15 wird.

Es wird eine Methode zur Optimierung der Web-Präsenz in Computernetzen zur Verfügung gestellt, durch welche das Benutzerverhalten erfasst und analysiert wird, wobei fehlerhafte Einflußfaktoren eliminiert werden und eine  
20 Optimierung der Web-Präsenz mit dem Ziel der Erhöhung der betriebswirtschaftlichen Wirksamkeit der Web-Präsenz erfolgt derart, daß aus den erfassten, bereinigten Benutzeraktivitäten und Aussagen aus einer Erfahrungsdatenbank ein Regelwert zur Realisierung eines  
25 geschlossenen Regelkreises gebildet wird, welcher Einfluß auf die Gestaltung der Web-Präsenz ausübt.

Die Erfindung soll nachstehend anhand von zumindest teilweise in den Figuren dargestellten Ausführungs-  
30 beispielen näher erläutert werden.

Es zeigen:

Fig. 1      Ablaufdiagramm des Verfahrens;

Fig. 2      Darstellung der am Verfahren beteiligten Komponenten;

Fig. 3 eine grafische Darstellung der Analysedaten.

Das erfindungsgemäße Verfahren erlaubt es einem Betreiber eines Servers für Internet-Domänen (Home-pages),  
5 alle Besuche, Aktivitäten und Umgebungsparameter der Besucher (User) dieser Domänen automatisch zu analysieren, diese Informationen mit anderen, in Datenbanken gespeicherten Daten zu vergleichen und auszuwerten und dadurch die Internetpräsenz seiner Domänen automatisch  
10 zu optimieren.

Das geht folgendermaßen vor sich:

Sämtliche Anfragen von Usern an den Server, auf dem die Homepage liegt, werden protokolliert. Diese Anfragen und die darauf folgenden Antworten des Servers werden  
15 im sogenannten Logfile zeitlich erfaßt. Diese Erfassung ist auf die technische Struktur des Servers ausgelegt, da die Serveraktivität protokolliert wird. Der Logfile wird von diversen herkömmlichen „Analyse“-Programmen als Basis für die Erfolgsmessung des Internets genutzt.  
20 Auf die Unrichtigkeit dieser Ergebnisse wurde bereits eingegangen. Bei der reinen Logfile-Analyse werden ausschließlich historische Daten erfaßt und dargestellt.

In den folgenden Abschnitten sollen die relevanten Vorgänge bei der Kommunikation zwischen Besucher und Server, auftretende Probleme und deren Lösung durch das Verfahren beschrieben werden.  
25

#### **HIT-Problematik:**

30 In der Anfangszeit des Controllings wurden einfach die Zeilen im Logfile gezählt. Jede Anfrage eines Besuchers löst einen sogenannten Hit aus. In der heutigen Zeit ist jedoch eine Hit-Zählung ungeeignet. Eine Bildschirmseite, die im HTML-Format kreiert worden ist,

kann neben dem Textkörper auch Multimediaelemente wie Bilder, Tondateien oder Video enthalten. Wird nun diese Seite angefordert, dann werden parallel die dazugehörigen Multimediaelemente geladen, die ebenfalls im  
5 Logfile protokolliert werden und einen Hit auslösen. Eine Seite mit vielen Multimediaelemente erzeugt also eine höhere Zugriffszahl und somit mehr Hits, als eine Seite, die rein auf Textbasis gestaltet ist.

In dem erfindungsgemäßen Verfahren wird der Hitzählung  
10 nur ein untergeordneter Status beigemessen, da dieser Wert nur einen Anhaltspunkt über den Trafic erlaubt.

In dem Verfahren werden deshalb weiterhin die Hits einer Domäne gezählt. Ihnen wird aber keine große Bedeutung beigemessen. Im Gegensatz zu den auf den gebräuch-  
15 lichen Analysemethoden beruhenden Zählungen werden die Hits allerdings genauer sein. Die genaueren Aussagen werden durch das temporäre Abschalten des Proxy-Cache-Servers und dem anschließenden Berechnen der nicht registrierten Zugriffe getroffen.

20

#### **Logfile-Problematik:**

Aufgrund der Verfälschungen durch fehlerhaft ausgewertete Hits kann also ein unkontrolliertes Erfassen der Hits nicht in Frage kommen. Es müssen also die Dinge  
25 erfaßt werden, die unabhängig von der grafischen Opulenz der Website sind. In diesem Fall bietet es sich an, die Anforderungen von einzelnen „Rahmendateien“ (HTML-Dokumente) zu zählen. Die Anforderung einer HTML-Seite wird in PageViews gezählt. Im Gegensatz zu  
30 PageViews werden bei dem vom Verfahren eingeführten PageImpression nur noch die Seiten gezählt, die als contentführend eingestuft werden.

Das Verfahren erlaubt die genaue Ermittlung von PageViews. Die Einheit "PageView" ist natürlich nicht re-

presentativ für eine Gesamtanalyse, gibt aber heute die Grundlage ab zur Berechnung des 1000er Kontaktpreises für Werbeflächen im Internet. Der Werbetreibende erhält durch diese Einheit den Eindruck, daß seine Werbebotschaft öfters angezeigt wurde, als es tatsächlich der Fall war.

Die Einheit "PageView" wird natürlich ebenfalls vom Verfahren gezählt. Hier treffen zwei Probleme aufeinander: zum einen werden zu viele Zugriffe gezählt, die das Verfahren in einem späteren Schritt korrigiert, zum anderen ruft die unten näher beschriebene Proxy-Cacheserver-Technik den gegenteiligen Effekt hervor: Seitenanforderungen werden aus Cache-Speichern bedient und führen zu keinem Eintrag im Logfile. Durch ein komplexes Verfahren wird jedoch sichergestellt, daß alle angeforderten HTML-Dokumente gezählt werden. Ein CGI-Image von nur wenigen BIT erzwingt einen Zugriff auf den Domäne-Server. Die eigentlichen Seiten können vom Cache-Server geladen werden. Mit dieser Methode bleibt eine gute Performance erhalten, es wird aber ein Hit auf dem Domäne-Server ausgelöst. Eine anschließende Berechnung ermittelt für diesen CGI-Impuls die nötige Anzahl von PageViews. In einem weiteren Schritt werden natürlich sämtliche Zugriffe von Suchmaschinen aus der Analyse der PageViews entfernt, denn diese sind keine Besucher im Sinne der Analyse. Als Ergebnis erhält der Betreiber einer Domäne eine 100% Aussage über die Anzahl der PageViews. Wie bereits erwähnt, ist dieser Wert nicht repräsentativ aber immer noch gebräuchlich in der Bestimmung der Werbeträgerleistung. Somit wird das Verfahren sämtliche Kennzahlen unterstützen und darüber hinaus genauer darstellen, auch wenn diese Werte keine oder nur geringe Aussagekraft haben.

**Frame-Problematik:**

Zwar wird heute PageView und PageImpression als Basis zur Berechnung des 1.000er Kontaktpreis angewendet, aber für ein effizientes Controlling sind diese Analysen immer noch nicht geeignet. Mit der Einführung der „Frame-Technologie“ besteht die Möglichkeit, eine Bildschirmseite aus mehreren HTML-Seiten aufzubauen. Beim Laden dieser Bildschirmseite wird jedoch jeder Frame mit einer dazugehörigen HTML-Datei und den ebenfalls verknüpften Bildern im Logfile protokolliert.

Je nach Frame-Struktur und grafische Opulenz einer Internetseite differiert der Quotient zwischen Anzahl von Zugriffen (Hit) und den gesehenen Bildschirmseiten (ContentView). Aus diesem Grund ist die Erfolgsmessung nach dieser Methode unzureichend.

Da sich im Laufe der Zeit die Programmierung in Frame-Technologie immer mehr durchgesetzt hat, ist die Lokalisierung der contentführenden Seiten zu Auswertungszwecken erstes Gebot und eins der entscheidenden Ziele des Verfahrens.

Natürlich entspricht jede erstellte Domäne bestimmten Erwartungshaltungen. Dies ist auch der Grund für die vielfältigen und unterschiedlichen Darstellungsformen von Internetseiten. Ein Seitenaufbau von 2 Frames oder ein Seitenaufbau von 7 Frames sind nichts ungewöhnliches und können sinnvoll sein. Das Verfahren berücksichtigt die individuellen Gestaltungsspielräume von Domäne-Betreibern. Das Verfahren ermöglicht das Auffinden von denjenigen Seiten, welche Definitionen über den Seitenaufbau beinhalten. Exakt diese Seiten beinhalten weder Navigation, werbeführende Botschaften noch Content-Seiten. Sie sind also ausschließlich für das technische Funktionieren der Seite verantwortlich und für die Beurteilung des Auftrittes uninteressant. Die-



ses Auffinden der „technischen“ HTML-Dokumente reduziert die Gesamtzahl der PageViews um ca. 30%. Die restlichen HTML-Dokumente werden Ihrer Bestimmung nach in einer Tabelle abgelegt und mit Attributen versehen, wie z. B. contentführend, Navigation oder werbeführend. Mit dieser Zuordnungs-Tabelle ist es möglich, aus dem Logfile alle HTML-Anforderungen ihrer Bestimmung nach korrekt zuzuordnen und zu zählen. In Kombination mit der schon oben beschriebenen Methode der CGI-Images ist so die exakte Anzahl von Zugriffen auf die einzelnen HTML-Dokumente sichergestellt. Somit ist erstmals die Ermittlung der ContentViews zu 100% möglich.

#### **CGI-Problematik:**

Eine weitere Herausforderung sind die dynamisch erzeugten Seiten. Hier wird aus Datenbanken entsprechend der Vorauswahl des Users die Seite serverseitig generiert und an den User zurückgeschickt. Aus diesem Grund wird kein Eintrag einer HTML-Seite im Logfile vorgenommen. Im Fall, daß dynamisch erzeugte Seiten aufgebaut werden, bzw. daß die Frame-Technik zum Einsatz kommt, sind alle bisher erläuterten Einheiten und Methoden für ein Controlling ungeeignet. Auf der anderen Seite sind gerade die so generierten Seiten für eine Produkt- und Finanzplanung hochinteressant, da sie in der Regel durch die Vorauswahl des Besuchers eine detaillierte Aussage über die Wünsche und Interessen des Besuchers bzw. Kunden enthalten.

Immer mehr HTML-Dokumente, gerade im e-commerce-Sektor, generieren Ihren Content dynamisch. D.h. in einem HTML-Rahmen werden Informationen aus einer Datenbank generiert und dargestellt. Aus diesem Grund ist es nicht direkt möglich aus dem Logfile einen Rückschluß zwischen dem aufgerufenen HTML Dokument und den dynamisch

generierten Informationen zu schließen. Das Verfahren ermöglicht aber neben dem Logfile ein weiteren eigen erzeugten File parallel zu erstellen, wo die Anforderungsparameter aus dem HTML-Dokument protokolliert werden. So ist es also möglich neben dem Rahmen HTML-Dokument auch die Informationsinhalte zu protokollieren und diese in einen Bezug zu setzen. Die Einheit ContentView wird, was auch den Tatsachen entspricht, um die Informationsinhalte von dynamisch erzeugten HTML-Dokumente quantitativ, also zahlenmäßig und qualitativ, also inhaltsbezogen, richtig erfaßt und dargestellt.

**Proxy-Cacheserver-Problematik:**

Proxy-Cacheserver dienen dazu, die vorhandene Bandbreite des Internets besser auszunutzen. Dieses wird durch den Einsatz von Cacheservern erreicht, indem diese Internetseiten zwischenspeichern. Fordert nun ein am Proxy-Server angeschlossener User eine Internetseite an, überprüft der Server ob die Seite bereits lokal vorliegt. Nur wenn die Seite nicht vorliegt, wird auf den Server des Seitenanbieters zugegriffen und ein Eintrag im Logfile vorgenommen. Dieser Eintrag enthält jedoch in den meisten Fällen die IP Adresse des Proxy-Cacheservers und nicht die des Users der die Seite angefordert hat. Im anderen Fall erhält der User die Seite aus dem Cacheserver, welches zur Folge hat, daß im Logfile des Seitenanbieters kein Eintrag erfolgt.

Zur genauen Auswertung ist es erforderlich, daß ein Seite die der User angefordert, serverseitig erfaßt und nicht unbemerkt und unregistriert vom Proxy-Cache an den User gesendet wird.

Die durch die Proxy-Cacheserver-Methode verbesserter Gesamt-Performance birgt auch einige Nachteile. Häufig aktualisierte Content-Seiten dürfen nicht in einem

Cache vorliegen, da sonst nicht sichergestellt werden kann, daß die darin enthaltenen Daten auch wirklich aktuell sind. Normalerweise wird in der Programmierung der Befehl „NOCACHE“, also „cachen nicht erlaubt“ eingegeben. Es gibt aber Provider, die diesen Befehl absichtlich unterdrücken, da ansonsten die Leistungsgrenze der Provider überschritten würde. Das Einfügen des Befehls „NOCACHE“ garantiert also nicht zu 100%, daß die Informationen direkt vom Domäne-Server abgerufen und so im Logfile protokolliert werden.

Dieser Problematik begegnet das Verfahren durch das Einbringen von erzwungenen Zugriffen auf den Domäne-Server bei content- bzw. werbe- führenden Seiten. Ein mittels CGI generierter Befehlssatz von nur 50 kb erzwingt bei jedem Browserzugriff auf eine Seite die Übertragung eines 1 Pixel großen Images vom Domäne-Server, auch wenn diese Seite bereits im lokalen Cache der Useres oder dem des Proxy-Servers vorliegt. Die Seite selbst unterliegt auf Wunsch weiterhin den Vorteilen des Cache-Mechanismus. Die Netzlast wird lediglich um jene 50 kb pro gezählter Seitenzugriff erhöht. Dieser Wert spielt in der heutigen Netzbandbreite keine Rolle und wirkt sich so nicht negativ auf die gesamte Performance der Seite aus.

25

**lokale Cachespeicher-Problematik:**

Ein vergleichbarer Effekt wird von dem lokalen Cachespeicher, der auf dem Rechner des Users vorliegt, ausgelöst. Alle bereits einmal übertragenen Seiten werden auf dem Rechner des Users im Cache-Verzeichnis abgespeichert. Dies führt zu einem beschleunigten Aufbau der Seite und wird heute von allen gängigen Browser unterstützt. Auch dieses Verfahren führt zu einer verrin-

30

gernden Zählung von Besuchen auf dem Server des Seitenanbieters.

Beispiel (Zeitverhalten): Ein User fordert die Seite 1 an und entschließt sich aufgrund eines Querverweises  
5 beim Lesen eine andere Seite aufzurufen. Nach kurzer Zeit wechselt er zurück auf Seite 1. Diese Seite wird nun aus dem lokalen Cache-Verzeichnis wiederhergestellt. Die Messung der Verweildauer im Logfile wird nun fälschlicherweise der Seite 2 zugeordnet unter der  
10 Prämisse, daß eine noch nicht im Cache vorhandene Seite vom Server angefordert wird. Hier ist zu beachten das die lokale Cachespeicher-Problematik mit der Proxy-Cacheserver-Problematik sich verstärkend auf die Fehlinterpretation auswirken kann.

15 Zur genauen Auswertung ist es erforderlich, daß ein Seite die der User angefordert, serverseitig erfaßt und nicht unbemerkt und unregistriert vom lokalen Browser-Cache an den User gesendet wird.

Die heute verwendeten Browser ermöglichen dem User, die  
20 Verwaltung des Cache-Speichers individuell einzustellen. Da es aus diesem Grund unzählige Variationen von Einstellungen gibt, kommt auch bei der lokalen Cache-Problematik das Verfahren welches bei der Diskussion der Proxy-Cacheserver-Problematik beschrieben  
25 ist, zum tragen. Ein mittels CGI generierter Befehlssatz von nur 50 kb erzwingt bei jedem Browserzugriff auf eine Seite die Übertragung eines 1 Pixel großen Images vom Domäne-Server, auch wenn diese Seite bereits im lokalen Cache der Useres vorliegt. Mit dem Verfahren  
30 wird neben der Proxy-Cache-Problematik auch die lokale Cache-Problematik gelöst.

***dynamische IP-Problematik:***

Eine weitere Herausforderung bei der Ermittlung von Besuchern stellt die Vergabe von dynamischen IP-Adressen dar. Viele Internet-Service-Anbieter vergeben sogenannte dynamische IP-Adressen. Diese IP-Adresse ist zwingend erforderlich in der Internet-Welt. Das Internet arbeitet nach dem Client-Server-Prinzip. Der Client stellt eine Anfrage an den Server. Der Server bearbeitet diese Anfrage und sendet das Ergebnis an die IP-Adresse der Anfrage zurück. Für die eindeutige Zuweisung der Nachricht ist also von beiden Seiten gesehen die IP-Adresse zwingend erforderlich. Die IP-Adresse ist in einem standardisierten Format beschrieben. (www.musterfirma.de oder 195.128.102.64). Es gibt nun weltweit mehr Internetnutzer als IP-Adressen. Aus diesem Grund müssen sich mehrere Internetnutzer wenige IP-Adressen teilen. Diese Methode müssen speziell große Dienstanbieter einsetzen. Hier bilden sich zwei Gruppen in der dynamischen IP-Vergabe.

Typ 1: Der User erhält bei der Einwahl in das Internet eine dynamisch IP-Adresse und die er für die gesamte Online-Session behält.

Typ 2: Hier erhält der User bei jeder Anforderung eines Bildschirmseiten-Elementes (Bild, Text, etc.) eine neue IP-Adresse. Bei starkem Datenverkehr beim Serviceanbieter wird bei einem erfolgreich beantworteten Element diese IP wieder freigegeben.

Sollten keine speziellen Zusatzverfahren angewendet werden, ist es also unmöglich die Anzahl der Besuche zu ermitteln. Auf dem letzten Internet-Kongreß wurde beschlossen, das Format um 1\*3 bzw. 2\*3 Stellen zu erweitern. Bis dieses weltweit in der Praxis eingeführt ist, werden wohl noch ca. 2 Jahre vergehen, in der die

Problematik durch den starken Zuwachs an Usern verstärkt wird.

Eine Bereinigung von möglichen Fehlinterpretationen von dynamischen IP-Adressen kann leider nur bedingt erreicht werden. Eine Möglichkeiten diese Problematik zu kompensieren, ist einen User unabhängig von seiner IP zu identifizieren. Ist dieser User nun festgestellt, dann kann er als einzelner Besucher erkannt werden, auch wenn er mit unterschiedlichen IP's innerhalb einer Session auftritt. Möchte der User sich nicht identifizieren, so besteht weiterhin die Möglichkeit über eine Zuordnungslogik die Wahrscheinlichkeit zu ermitteln, daß die unterschiedlichen IP's einem Besucher zugeordnet werden. Die Darstellung der Besucherzahlen erfolgt dann nach den Gesetzen der Aussagenwahrscheinlichkeit in Bandbreiten. Es wird also neben den tatsächlich vorhandenen Besucherzahlen eine weitere Grafik geben mit dem Höchstwert an Besuchern, dem niedrigsten Wert an Besuchern und der berechneten Anzahl von Besuchern.

Leider muß vorausgeschickt werden, daß die Problematik nicht 100%ig gelöst werden kann. Die Vorteile des Verfahrens beruhen jedoch auf der Erkenntnis, zwischen sicheren und unsicheren Daten zu unterscheiden, diese zu filtern und diese Daten in einer Bandbreite zu präsentieren. Durch das Verfahren kann diese Problematik sogar bis zu 98% Aussagengenauigkeit gelöst werden. Somit sind die Ergebnisse des Verfahrens so genau, daß normgerechte Planungsdaten berechnet werden können. 2% Abweichung vom realen Wert spielen für die strategische Weiterentwicklung keine entscheidende Rolle. Der Lösungsansatz beruht auf zwei grundsätzlich unterschiedlichen Userverhaltensmustern.

Fall 1: Der User identifiziert sich (wahlweise auch unter einem Nickname)

Fall 2: Der User möchte auf jeden Fall unerkantnt bleiben

Betrachten wir nun Fall 1: In diesem Fall hat der User  
5 die Cookies in seinem Browser aktiviert. In diesem Fall  
wird eine Variable in das Cookie eingetragen. Sollte  
nun der Besucher die Homepage besuchen, kann er während  
der Session anhand seines Cookieeintrages identifiziert  
werden. Besucht er nun zu einem späteren Zeitpunkt die  
10 Domäne erneut, so wird diesem Besucher ein erneuter Be-  
such zugewiesen, selbst wenn er unter einer neuen IP-  
Adresse erscheint. Mit dieser Methode (unter der Prä-  
misse das 100% der User Cookies aktiviert haben) kann  
eine eindeutige Aussage, die wie folgt aussehen könnte,  
15 getroffen werden.

„Im Monat September des Jahres 1999 hatten Sie auf  
Ihrer Homepage 650 Besuche. Diese 650 Besuche wur-  
den von 120 unterschiedlichen Besuchern ausgelöst.  
Sie haben im letzten Monat 40 neue Besucher gewon-  
20 nen. ... usw.“

Die Erfahrung zeigt aber, daß nur 50% der User Cookies  
aktiviert haben. In der Konsequenz bedeutet dieses nun,  
daß eine unbekannte Masse von 50% übrig bleibt, die  
25 nicht zweifelsfrei identifiziert werden kann. Eine Mög-  
lichkeit, diese Masse trotz unterschiedlicher IP's ein-  
zelnen bestimmten Besuchern zuzuordnen, ist die Verfol-  
gung, welche Seiten nacheinander aufgerufen werden, der  
sogenannte Clickstream. Hilfreich hierfür sind weitere  
30 Informationen, die über den normalen Logfile nicht pro-  
tokolliert werden. Ein weiterer Befehl für die Zuord-  
nung unterschiedlicher IP's auf einen Besucher ist  
„VIA“. Via enthält die gesamte Kette der Proxy-  
Weiterleitungen. Anhand der Weiterleitung, sprich des

Weges, und dem logischen Zusammenhang im Clickstream des Logfiles kann nach der Wahrscheinlichkeitsrechnung eine extrem hohe Aussagengenauigkeit getroffen werden. Die gesamte Sicherheit dieses Teilverfahrens beläuft sich auf mehr als 98% und ist daher als sehr sicher einzustufen.

**Router-Problematik:**

Viele Unternehmen setzen aus Sicherheitsgründen Firewall-Rechner ein. Ein Firewall-Rechner dient der Sicherheit im Unternehmen. Es wird ein „Nadelöhr“ geschaffen, durch das alle Informationen nach innen und nach außen fließen müssen. Dieses Nadelöhr ist wesentlich leichter zu überwachen als viele offene Schnittstellen nach außen. Dieser Firewall setzt sämtliche im Unternehmen verwendete IP auf eine einzige externe IP-Adresse um. Im Falle, daß dieses Unternehmen keine Proxy-Cache-Server verwendet, erscheint grundsätzlich nur diese Firewall-IP im Logfile. Das heißt, auch verschiedene User im Unternehmen erscheinen unter der gleichen IP-Adresse und machen eine Differenzierung ohne zusätzliche Verfahren unmöglich.

Beispiel: 3 Besucher eines Unternehmens besuchen gleichzeitig eine Internet-Domäne. Durch die Vergabe einer IP an alle drei User erscheint der Eindruck, das eine Person sehr schnell ungeordnet durch viele Kapitel und Unterkapitel springt.

Verwendet das Unternehmen jedoch ein Proxy-Cache-System, welches ja empfehlenswert ist, ist die Situation ähnlich wie bei den großen Internet-Service-Anbietern.



Dies ist eine ähnliche Herausforderung wie die Lösung der dynamischen IP. Eine 100% Aussage kann nur getroffen werden, wenn der User sich identifizieren läßt. Sollte dies nicht der Fall sein, so kann nur nach statistischen Modellen ermittelt werden, wie hoch die Wahrscheinlichkeit ist, daß ein bestimmter User die Anfragen an den Server stellt. Untersuchungen und Tests, die durchgeführt wurden, zeigen, daß sich ca. 50% aller User identifizieren lassen, so daß eine hohen Aussagegenauigkeit über die Anzahl der restlichen Besucher getroffen werden kann. Mit diesen Lösungsansätzen kann eine bis zu 98% genaue Aussage über die Anzahl der Besucher getroffen werden.

Die Router- bzw. Firewall-Problematik läßt sich mit der gleichen Methode, wie im Falle dynamischer IP's beschrieben, lösen (Clickstream, Verfolgen der Proxy-Weiterleitungen). Wobei jedoch anzumerken ist, daß die Aussagengenauigkeit beim Auftreten dieser Problematik etwas niedriger ist. Sie liegt bei ca. 92%. Dieser Wert fließt jedoch nur zu einem geringen Maße und branchenabhängig in die gesamte Auswertung ein und sollte individuell je nach dem Informationsbedürfnis berücksichtigt werden.

Eine weitere Möglichkeit zur Identifizierung kann, unter der Voraussetzung das PHP auf dem Server installiert ist, durch „Sessiontracking“ erreicht werden. Alle Dokumenttypen werden dann von \*.htm bzw. \*.html in \*.php umbenannt. Beim Aufruf der Startseite der Domain (meist Index.php) wird eine eindeutige Variable dem User zugeteilt. Diese Variable bleibt dem über die ganze Session konstant und wird an alle aufgerufenen Dokumente der Session angehängen. Diese Anhang ist im Standard- Logfile sichtbar und für die Auswertung selektierbar. Sessiontracking ist mit Sicherheit der „Königsweg“. In

Kombination mit Aktivierung der Cookies oder freigeschaltete Pentium III ID ist eine nahezu 100% Identifizierung des Userverhaltens möglich. Ebenfalls kann zweifelsfrei festgestellt werden, wie oft ein User die  
5 Seite besucht hat. (Differenzierung Neu-/Altkunde)

**Robot-Problematic:**

In regelmäßigen Abständen besuchen Suchmaschinen die Internetseiten, um nach Neuerungen und Aktualisierungen  
10 zu suchen. Diese Besuche geschehen vollautomatisch durch sogenannte Suchrobots. Suchrobots sind die kleine Softwarespione, die das Netz absuchen und sämtliche Neuerungen in das Datenbanksystem des Suchmaschinenbetreibers schreiben. Diese Besuche lösen natürlich  
15 ebenfalls Zugriffe im Serverlog aus, die fälschlicherweise als Besuch gewertet werden könnten. Diese Visits sind mit Sicherheit am einfachsten heraus zu filtern, sollten aber nicht unbeachtet bleiben.

Ziel muß es sein, die Besuche von Suchmaschinen auf einer Domäne aus dem Logile zu filtern, so daß diese  
20 nicht die Auswertungsdaten verfälschen. Weiterhin sollen diese Besuche in einer separaten Auswertung dargestellt werden, um Informationen über die verwendeten Suchmaschinen mit den dazugehörigen Suchworten zu erhalten.  
25

Fast alle gängigen Suchmaschinen greifen auf die Datei „ROBOT.TXT“ zu. Ein User wird niemals auf diese Datei in der Verzeichnisstruktur zugreifen. Aus diesem Grund kann im Logfile der Besuch von Suchmaschinen gefiltert  
30 werden und von den Userdaten abgezogen werden. Der Besuch von Suchmaschinen kann nun in eine separate Auswertung dargestellt werden. Aus dieser Auswertung können nun weitere Entwicklungsstufen für einen optimierten Metatag eingeleitet werden.

**Suchmaschinen- und Link-Problematik bei frame-orientierten Seiten:**

Diese Problematik wird durch mangelndes Wissen der Seitenhersteller über die Internetfunktionsweise ausgelöst. Wenn eine Suchbegriff in eine Suchmaschine eingegeben wird, und die Suche erfolgreich war, werden die Ergebnisse der Suche dem Anfragenden in der Ergebnisliste angezeigt. Klickt dieser nun auf die angegebene URL, dann erscheint meistens nur die Contentseite ohne die dazugehörige Framestruktur.

Leider findet man auf den meisten Seiten, die frame-orientiert aufgebaut sind, diesen Fehler. Verstärkt wird dieser Fehler durch den Komfort und die Leistungsfähigkeit moderner Suchmaschinen. Wenn eine URL bei einer Suchmaschinen angemeldet wird, dann werden die Suchroboter diese URL innerhalb der nächsten Tage besuchen. Mit verschiedenen Parametern mit METATAGS beziehungsweise Parametern im ROBOT.TXT wird es zugelassen bzw. verhindert, daß bestimmte Unterverzeichnisse durchsucht werden. Wird das Durchsuchen von Unterverzeichnissen zugelassen, was ja auch sinnvoll ist, muß dafür Sorge getragen werden, daß die komplette Framestruktur im Falle eines Userzugriffes über die Suchmaschine mit aufgebaut wird, um die Navigation zu ermöglichen. Nur so ist es möglich die komplette Internetdomäne dem User zu präsentieren. Diese Programmierung wird meistens vergessen.

Aus einschlägigen Untersuchungen ist bekannt, daß 40% aller Ergebnisse von Suchmaschinen in eine Unterverzweigung der Homepage führen. Diese große Zahl von „Quereinsteigern“ muß ordnungsgemäß bedient werden. Eine programmiertechnische Überarbeitung, gerade der von Laien erstellter Homepages ist unumgänglich, da diese

Seiten auch noch x-fach in der Ergebnisliste der Suchmaschine angezeigt werden.

Ziel muß es sein, daß beim direkten Aufruf einer Content-Seite die gesamte Frame-Struktur aufgebaut und der User zu dieser Seite durchgeschleift wird.

Die Erfahrung zeigt, daß bei einer nicht unerheblichen Anzahl von Domänen gravierende Fehler im Quellcode vorliegen, bzw. Metatags falsch oder gar nicht generiert worden sind. Ein weiterer potentiell vorliegender Fehler, gerade bei frame-orientierten Internetstrukturen, ist der falsche Aufbau der Seite, wenn sie in einem untergeordneten Content aufgerufen wird und nicht von der Homepage aus. In diesem Fall wird die Navigation nicht mehr angezeigt. Diesen Problematiken wird bei dem Verfahren mit den geeigneten Methoden begegnet, die automatisch bzw. halbautomatisch den Code auf geeigneter Weise optimieren und korrigieren.

Die oben beschriebene Maßnahme ist für das Gesamtverfahren von großer Bedeutung. Zum ersten wird der Qualitätssicherung genüge getragen und zum zweiten liegen die Daten des Metatags und somit die Informationen für die Suchmaschine in geeigneter Form in einer Datenbank vor. So ist es erstmals möglich einen Soll/Ist-Vergleich zwischen den meisten gesuchten Schlagworten, den Suchmaschinen Fabrikaten und den tatsächlich vorliegenden Informationen auf einer Internetseite durchzuführen. Die Ergebnisse dieser Vergleichsstudie ermöglichen die gezielte Optimierung der Metatags für kundenorientierte Suchergebnisse und somit für den Erfolg einer Domäne.

30

#### **Interpretations-Problematik:**

Die einzige Möglichkeit Daten über das Userverhalten (es sei denn, der User identifiziert sich) zu erhalten, ist der Logfile. Der Logfile ist aber in seiner Funkti-

on für die Protokollierung der Servertätigkeit ausgerichtet und nicht unbedingt für weitere Interpretationen bestimmt. Einige Hersteller von Interpretationssoftware lesen die Daten aus und geben diese ungefiltert und nicht an die Bedürfnisse angepaßt an diverse Analysetools weiter. Fehlende Informationen über das Userverhalten und die technische Performance werden einfach angenommen bzw. übernommen. Die oben genannten Problematiken bleiben zum großen Teil unberücksichtigt. Die Konsequenz sind pseudogenaue Basisdaten, die weder für das Controlling noch für die Planung akzeptabel sind. Im Gegenteil: Es werden sogar falsche Daten als Bemessungsgrundlage aufgeführt, die eine Planung in die falsche Richtung verursachen und somit Schaden anrichten.

Ziel ist es, daß der Seitenanbieter zu Planungs- und Beurteilungszwecken nur die Daten erhält, die valide und aussagekräftig sind. Die Daten, die ausschließlich auf Vermutung basieren, sind in der Analyse zu eliminieren, so daß diese die Auswertung nicht verfälschen können.

Beispiel: Einige Analyse-Programme treffen die Aussage, daß 99% der User Cookies verwenden. Diese Aussage basiert auf dem Prozentsatz der cookiefähigen Browser. In Wirklichkeit haben ca. 50% bis 60% der User „Cookies akzeptieren“ ausgeschaltet, so daß der Wert „99% der User haben Cookies“ falsch ist. Solche Zahlen können zu dramatischen Fehlentscheidungen im e-commerce-Sektor führen. Ein Unternehmen, welches sein Shop-System cookiebasiert aufbaut, beschneidet seinen Kundenkreis aufgrund der Falschaussagen um 50%-60%. Sollte dieses bereits in den Brunnen gefallene Kind zu spät

erkannt werden, sind Umsatzeinbußen zu ver-  
zeichnen, die mit einem normgerechten Tool  
nicht entstanden wären. Eine Neuprogrammierung des Shopsystems unter Java, CGI, etc.  
5 ist unumgänglich.

Unsichere Daten bzw. empirisch ermittelte Daten sind  
auf Ihre Aussagengenauigkeit und Wahrscheinlichkeit zu  
prüfen und in den daraus resultierenden Toleranz-  
10 bereichen darzustellen.

Die hier dargestellten Herausforderungen werden durch  
die einzelnen Module des Verfahrens kompensiert. Die  
Erfinder des Verfahrens maßen sich nicht an, alle Daten  
mit einer 100% Sicherheit zu erhalten. Bei diesem Ver-  
15 fahren wird aber unterschieden zwischen Aussagen, die  
100% sicher und geprüft sind, und den Daten, die mit  
einer Ungenauigkeit behaftet sind. Diese unsicheren Da-  
ten werden nach anerkannten statistischen Methoden auf  
Ihre Aussagewahrscheinlichkeit geprüft und in einem To-  
20 leranzfeld dargestellt.

Der User kann zwischen drei Varianten der Darstellung  
wählen.

1. Darstellen von den Daten, die sicher sind, und ge-  
trennt davon
- 25 2. Darstellung der Daten, die in einem Toleranzfeld  
liegen, sowie
3. die gemischte und gewichtete Darstellung beider  
Daten mit ebenfalls enthaltenen Toleranzfeld.

Auf diese Weise kann sichergestellt werden, daß keine  
30 Daten fehlerhaft interpretiert werden und so Fehl-  
entscheidungen weitestgehend ausgeschlossen werden kön-  
nen.

**Format-Problematik:**

Zur Zeit sind mehr als 30 verschiedene Logfile-Formate im Einsatz. Je nach Betriebssystem, Serverarchitektur und Provider kommen nicht nur die 30 verschiedene Basis-Formate vor, sondern auch noch die unterschiedlichsten Kombinationen. Eine weitere Problematik ist die Verwaltung dieser Daten. Viele Provider löschen historische Daten oder komprimieren diese und legen sie in entsprechenden Unterverzeichnissen ab. In diesem Fall ist natürlich das Wissen des Internet-Administrators über den Umgang mit den Logfiles von großer Bedeutung. Von unabhängiger Entscheidungsfindung der Geschäftsleitung kann keine Rede mehr sein. Diese Vielzahl an Basis-Formaten des Logfiles, unterschiedlichen Kombinationen und eine uneinheitliche Verwaltung der historischen Daten auf Providerseite machen bisher ein einheitliche, lückenlose und vergleichbare Performance-Messung der Internet-Aktivitäten nahezu unmöglich. Das Ziel, die Format-Problematik zu kompensieren, kann nur erreicht werden, wenn die Daten in einen eigenen und unabhängigen Datenbanksystem protokolliert werden. Im besonderen ist zu berücksichtigen, daß herkömmlicherweise viele Provider die historischen Daten am Monatsende mit dem neuen Monat überschreiben. Eine langfristige Planung und freigestaltbare Analysezeiträume sind somit nur in den seltensten Fällen möglich.

Von der Basis her werden durch das Verfahren die unterschiedlichen Formate komplett unterstützt. Zu Beginn des Einsatzes des Verfahrens wird unter anderem ein Logfile downloadet und analysiert. Aus dieser Analyse wird der Informationsgehalt ermittelt und mit dem Anforderungsprofil des Controllers verglichen. Spezielle Zusatzinformationen wie z.B. Bildschirmauflösungen, „Cookies aktiv/nicht aktiv“ und die Ermittlung von Neu-/Altkunde sowie die Beseitigung von Proxy-Cache-

Problematiken werden in einer Erweiterung des Logfiles abgespeichert. Die wohl größte Herausforderung, die Beseitigung der unterschiedlichen Handhabung von Logfiles auf Providerseite wird durch das automatische Downloaden des Logfiles behoben. Die bereinigten Ergebnisse werden anschließend in ein eigens erstelltes Datenbanksystem lokal auf dem Rechner des Controllers abgespeichert. Aus dieser Datenmanagement-System werden alle weiteren Darstellungen generiert. Mit dieser Methode ist es möglich, verschiedene Zeitintervalle für die Analyse auszuwählen, bzw. verschiedene Zeitintervalle gegenüberzustellen.

Dieses Verfahren bietet somit einige entscheidende Vorteile:

- Die Datenbank ist unabhängig vom Logfile;
- Die Datenbank kann logisch mit anderen Datenbanken verknüpft werden (Soll/Ist-Vergleich);
- Es können individuelle Zusatzinformationen (z.B. Aktualisierungen) gespeichert werden;
- Unabhängigkeit vom Handling des Providers ist gesichert;
- Schnittstellen zu anderen Applikationen (z.B. zu Warenwirtschaftssystemen) werden bereitgestellt,
- Beliebige Bestimmung der Analysezeiträume wird gestattet;
- Es stellt eine Basis dar für Erweiterungen und Flexibilität bei nachfolgenden Ausbaustufen (mail-flow, comm-flow, etc.).

#### **Darstellungs-Problematik:**

Im Logfile werden alle Protokolldaten chronologisch abgespeichert. Um diese Daten „lesbar“ zu gestalten, müssen diese gefiltert, sortiert und interpretiert werden. Dieser Aufwand wird selten oder nur mit hohen Kosten



betreiben. Es kann nicht ohne weiteres ein Rückschluß von der technischen Verzeichnisstruktur, die im Logfile erkennbar ist, und der Präsentationsstruktur gezogen werden. Ebenfalls werden alle Dateien, auch die für die  
5 Navigation bestimmten Daten, unstrukturiert dargestellt. Eine weitere Herausforderung ist es, einen Bezug zwischen den Dateinamen und den präsentierten Inhalten zu finden.

Die Folge ist: Ausschließlich der Personenkreis, der  
10 die Internetseite entwickelt hat, ist in der Lage den Logfile und die Auswertung zu interpretieren. Es ist aber ein großer Aufwand für die Interpretation der Daten erforderlich, der in den seltensten Fällen betrieben wird.

15 Die Vertriebs- und Marketingabteilung hat keine Möglichkeit, unabhängig von der Entwicklungsabteilung der Internetseite, strategische Entscheidungen über die Zukunft der Domäne zu treffen.

Die Darstellung muß deshalb in der Form aufgearbeitet  
20 sein, daß ein Personenkreis ohne herausragende Kenntnisse im Internet, die Ergebnisse verstehen kann. Die Basis ist die klare Strukturierung des Contents, die unabhängig von der technischen Strukturierung ist. Ebenso ist die Darstellungsform "Text" als Träger der  
25 Information nicht zu unterschätzen und sollte neben der grafischen Darstellung ergänzend wirken. Frei gestaltbare Bemessungszeiträume und eine individuelle Reduzierung der Daten auf das Informationsbedürfnis des Interessierten werden deshalb vom Verfahren zur Verfügung  
30 gestellt.

Voraussetzung einer leichtverständlichen Darstellung ist die exakte Abbildung der Darstellungs-Struktur einer Internetseite. D.h. die Struktur muß dem entsprechen, was sich dem User auf dem Monitor zeigt. Diese

Struktur entspricht in den seltensten Fällen der technischen Struktur auf dem Server. Aus diesem Grund ist es zwingend erforderlich, eine „Übersetzung“ zwischen der technischen und der Darstellungsstruktur zu schaffen.

5 Durch diese „Übersetzung“ werden auch die Bezeichnungen der Dateien in eine für den Laien verständlichen Sprache gebracht. Wie bereits erwähnt steht zum Anfang des Verfahrens das Downloaden der kompletten Internet-Struktur als Offline-Version. Das Verfahren filtert nun alle

10 Bilder aus, da diese für das Userverhalten unerheblich sind. Ebenfalls unerheblich sind alle HTML-Seiten, die den Rahmen für den Content bilden, d.h. einen Frameset aufrufen. Die nun übriggebliebenen Dateien entsprechen nur noch folgenden Kriterien: Navigation, Content, Werbeflächen, Ausführende Seite, Download.

15 Die Dateien werden nun vom Controller den Punkten Navigation, Content und Werbefläche zugeordnet, indem Sie nacheinander angezeigt werden. Nach dieser Prozedur stehen die contentführenden Seiten fest und werden in

20 Kapitel und Unterkapitel strukturiert. Diese Informationen werden ebenfalls in dem Datenbanksystem abgespeichert und dient als Basis für die Auswertung und Darstellung der Analyse.

Ein weiterführende Abfragesystematik analysiert die individuellen Informationsbedürfnisse des Controllers.

25 Diese Daten werden ebenfalls im Datenbanksystem gespeichert. Mit diesem Verfahren wird ein starker Visualisierungsgrad erreicht, der den geforderten Ansprüchen dienen und das wesentliche erfaßt.

30 Das Verfahren umfaßt die klassische Logfile-Analyse zur Dokumentation der chronologischen Abläufe auf Serverseite, wie die Dokumentation der Hits sowie der Errors der einzelnen HTML-Seiten im zeitlichen Kontext, die

Darstellung von diversen Seitenzugriffsstatistiken in Form von Balken- und Kreisdiagrammen und die Ermittlung des vom Besucher benutzten Browser-Typs. Erweitert wird die klassische Analyse durch ein Web-Analyse-System für  
5 eine aussagekräftige Dokumentation des tatsächlichen User-Verhaltens im Internet. Dieses Web-Analyse-System umfaßt die programmtechnische Optimierung des Internet-Auftrittes, die Darstellung fehlerbereinigter Informationen über das User-Verhalten, wie z. B.: content  
10 views, erfolgte Downloads, Pageviews; die Werbeträgerleistung nach IVW-Methode, die Darstellung des exakten zeitlichen User-Verhaltens, Informationen über den User wie verwendetes Betriebssystem, Browser, Farbtiefe, Auflösung des Monitors, tatsächliche Cookie-  
15 Aktivierung, JAVA-Fähigkeit, VB-Fähigkeit; Darstellung der Traceability des Users: verwendete Suchmaschinen und Suchworte, Herkunft des Users über Links, Direktewahl oder Einwahl über Subdomain, Analyse der typischen Userpfade: Ein- und Ausstiegsseite, Verweilzeiten, Alt-/Neukunde; die Analyse der Verweilzeiten  
20 sowie ein Alarmsystem zur Analyse gering frequentierter Seiten. Ein weiterer Vorteil des Verfahrens gegenüber herkömmlichen Verfahren ist darin zu sehen, daß mit dem erfindungsgemäßen Verfahren ein Tool zur Analyse beliebiger Zeiträume unter Berücksichtigung des relativen  
25 und des absoluten Erfolges zur Verfügung steht. Durch den Zugang zu einer externen Erfahrungsdatenbank, in der das User-Verhalten dokumentiert wird, sind aussagekräftige Trendanalysen zukünftiger Entwicklungen und  
30 eine Evaluierung der User-Akzeptanz möglich.

Ein weiterer Teil des Verfahrens ist das Tool zur Optimierung bestehender Web Sites, welches automatisch eine

Code-Optimierung und Frame-Korrektur sowie Suchwort- und Meta Tag-Optimierung durchführt. Ebenso ist ein Tool für die Aktualisierung der Suchmaschinen-Einträge und der Eintragung selbst vorhanden. Ein weiterer Bestandteil wird in einem Trainings-Tool zur Verfügung gestellt. Dieser erlaubt ein web-basiertes Training, die Analyse zur Lernerfolgsmessung und die Evaluierung der Effizienz des Schulungsprogrammes. Ein Schulungsprogramm für effektives Web-Controlling unter Berücksichtigung rechtlicher Aspekte und technischer Voraussetzungen vermittelt die Darstellung der Basics sowie der Standardmethoden und bietet ein strategisch aufgebautes Controllingsystem zur Optimierung der interaktiven Kommunikationskette. Ergänzt wird das Verfahren durch eine Erweiterung um die typischen Funktionalitäten des e-commerce (PAP-Programmierung, CGI-Programmierung)

Die eben beschriebene Funktionalität des Verfahrens wird ergänzt durch die Möglichkeit, basierend auf der Datenanalyse Zielvorgaben-Kontrollfunktionen und Alarmsysteme zu etablieren, die ein effizientes Controlling ermöglichen, denn diese Daten sind für die strategische Weiterentwicklung und Optimierung der Internet-Domain unter dem Aspekt der Kundenorientierung von höchster Relevanz. Es wird damit ein Link zwischen der Analyse und der Optimierung geschaffen, um einen kontinuierlichen Verbesserungsprozeß (KVP) zu realisieren.

Weitere Funktionalitäten, die vom Verfahren bereitgestellt werden, bestehen in der Verifizierung der Werbewirksamkeit von Marketingmaßnahmen durch spezielle Analyse der Web Site, der Verifizierung der Response-Aktivitäten des Unternehmens auf Anfragen über die In-

ternet-Domain und einer Beurteilung der kommunikativen Aktivitäten auch außerhalb des Webs.

Das Verfahren in seiner Gesamtheit stellt somit einen  
5 Regelkreis dar, der am Beispiel eines Internet-  
Auftritts mit einer Homepage wie folgt beschrieben werden kann. Ausgehend von der Präsentation der Internet-Seite erfolgt im Anschluß die Erfassung und Analyse der  
Benutzeraktivitäten, die von Erfassungs- und Interpre-  
10 tationsfehlern bereinigt werden. Über den anschließenden Vergleich mit von einer Erfahrungsdatenbank zur Verfügung gestellten Informationenerfolgt erfolgt sodann die Optimierung und Anpassung der Internet-Seite, mit dem Ziel die Vorgaben der Zielvorgaben-  
15 Kontrollfunktionen zu erreichen.

Detailliert kann der Prozeß folgendermaßen beschrieben werden:

Sämtliche Anfragen von Usern an einen Server werden  
20 protokolliert. Diese Anfragen und die darauf folgenden Antworten des Servers werden im sogenannten Logfile zeitlich erfaßt. Diese Erfassung ist auf die technische Struktur des Servers ausgelegt, da die Serveraktivität protokolliert wird. Der Logfile wird von diversen Pro-  
25 grammen als Basis für die Erfolgsmessung des Internets genutzt. Auf die Unrichtigkeit dieser Ergebnisse wurde bereits ausführlich eingegangen. Bei der reinen Logfile-Analyse werden ausschließlich historische Daten  
30 erfaßt und dargestellt. Zukunftsorientierte Informationen können nicht berechnet werden, da bereits die Eingangsinformationen mit Fehlern behaftet sind.

Folgende Informationen werden im klassischen Logfile dargestellt :

- Hits:** Darstellung aller Zugriffe incl. Laden von Bildern und Aufbau der Frame-orientierten Seiten;
- PageViews:** Darstellung aller HTML-Seiten - das Laden von Bildern wird nicht mitgezählt. Dieser Wert  
5 wird heute als Größe zur Berechnung der Werbeträgerleistung herangezogen;
- Errors:** Fehler, wenn Seiten angefragt werden, die nicht auf dem Server vorhanden sind bzw. nicht gefunden werden können;  
10
- Übertragene Bytes:** Menge von Daten, die vom Server übertragen worden sind;
- Tagesverhalten:** Zeitliche Darstellung der Informationsanfragen; Angabe der Uhrzeit, wann Informationen abgefragt wurden  
15
- Wochenverhalten:** Zeitliche Darstellung der Informationsanfragen in der Tagesübersicht;
- Monatsverhalten:** Zeitliche Darstellung der Informationsanfragen in der Wochenübersicht;
- Allgemeines zeitliches Verhalten:** Verweildauer auf einer Seite, mit dem Hintergrund der Intensitätsprüfung und Prüfung der Kommunikationsstärke der Seite;  
20
- Betriebssystem:** Zahlenmäßige Erfassung, mit welchem Betriebssystem der User arbeitet;  
25
- Browser:** Zahlenmäßige Erfassung, mit welchem Browser der User arbeitet;
- Suchmaschinen:** Statistische Erhebung, mit welcher Suchmaschine die Domain gefunden wurde;
- Suchworte:** Statistische Erhebung, mit welchen Suchworten die Domain gefunden wurde;  
30
- Herkunft des Users:** Erfassung des Landes, aus dem der Besucher der Domain kommt;

**User-Pfade:** Erfassung der Klickfolge innerhalb der Domain;

**Ein-/Ausstiegsseite:** Erfassung der Ein-/Ausstiegsseite Users innerhalb einer Domain

5

Wie bereits dargelegt, reichen die o.g. Daten zur Realisierung eines effizienten Controllings jedoch nicht aus. Basierend auf der klassischen Logfile-Analyse bietet das Verfahren folgenden Zusatznutzen:

10

- Transport der technisch orientierten Information in die reale Darstellungsform der Domain
- Kontrolle, ob die Daten richtig erfaßt wurden
- Bereinigung von technisch bedingten Fehlinterpretationen
- Empirische Ermittlung von unsicheren Informationen mit der Darstellung in Toleranz-werten
- Darstellung der Daten entsprechend der Bedürfnisstruktur des Controllers

20

Die Bedürfnisstruktur läßt sich lt. neuesten Untersuchungen wie folgt darstellen :

25

- Was ist unter zeitlichen Gesichtspunkten auf der Domain geschehen?
- Wie ist es um die Leistungsfähigkeit der Rechner-Performance der User gestellt?
- Wie hat der User die Domain gefunden, wonach hat er gesucht und woher kommt er?
- Was war die Intention des Users, hat er die Domain bereits vorher besucht und wie lange hat er welche Informationen abgefragt?
- Administrative Zusatzinformationen zur Überprüfung der Kommunikation. Wurden alle Mails beantwortet,

30

sind die Zielvorgaben erreicht worden, welchen Einfluß haben Anzeigen und Werbung z.B. in Printmedien auf die Zugriffe der Domain etc.

- 5 Durch das Verfahren können weitere für den Controller interessante Informationen generiert werden. Hier ist eine klare Erweiterung der Leistungsfähigkeit gegenüber den bisher bekannten Produkten zu sehen. Und zwar zeichnet sich das R.T.F.I. Verfahren durch folgende Zusatzfunktionen aus :
- 10

**ContentViews:** Anzahl von Seitenkontakten, die als contentführend bezeichnet werden können, also tatsächlich relevanten Inhalt haben, und nicht der Navigation dienen. Ein weiterer sekundärer Nebeneffekt ist die Analyse von Bannerwerbung und die daraus resultierende Abrechnung der Werbekosten. Diese werden zur Zeit aus den nicht korrekt erfaßten PageViews ermittelt;

15

20 **Seiten Statistik:** Wie verteilt sich die Besucherhäufigkeit auf den einzelnen Seiten zum gesamten Content? Wie sieht die Seite aus und wie lange wurde sie im Durchschnitt besucht?

**Downloads:** Welche und wie oft wurden bestimmte Dateien downloadet?

25

**Infoanforderung:** Wie oft wurde die Seite mit einer Infoanforderung besucht und wie oft wurde anschließend die Information tatsächlich angefordert?

30 **Aktualisierung:** Wann und was wurde an einer bestimmten Seite verändert bzw. aktualisiert?

**Farbtiefe:** Welche Farbtiefe hat der User auf seinem Rechner eingestellt?



- Auflösung:** Welche Auflösung hat der User auf seinem Rechner eingestellt?
- Prozessor:** Welchen Prozessortyp besitzt der User?
- Cookies:** Hat der User auch tatsächlich die Cookies aktiviert oder besitzt er nur einen cookiefähigen Browser?
- 5
- Java:** Welche Version von Java kann auf dem Browser des Users ausgeführt werden?
- VB:** Ist die Performance des Users in der Lage VB auszuführen?
- 10
- Links:** Wie werden die Verlinkungen genutzt? Die Messung erfolgt bidirektional, d.h. es werden abgehende und eingehende Links gemessen und verglichen.
- 15
- Direkteinwahl:** Wie viele User wählen die Domain direkt an und wie ist die prozentuale Verteilung zu den anderen Einwahlmöglichkeiten?
- Sub-Domain:** Wie viele User kommen über eine evtl. untergeordnete Domain auf die Haupt-Domain? Diese Möglichkeit wird häufig genutzt, um den Erfolg einer Aktion zu bewerten.
- 20
- Neu-/Altkunde:** War der Besucher schon einmal auf der Domain, wie oft kommt er in welchen Zeitintervallen erneut vorbei? Welches Profil besitzt er und was sind seine Informationsbedürfnisse ?
- 25
- Verweilzeiten:** Welche Benutzergruppe verweilt wie lange bei welchen Informationen? Wie lange bleibt der Besucher auf bestimmten Seiten und wie lange dauert sein gesamter Besuch?
- 30
- Anzahl der Besuche:** Wie viele Besuche hatte die Domain und von wie vielen Besuchern wurden die Visits ausgelöst?
- Aktionen:** Die Aktionsauswertung ist ein wesentlicher Faktor des Verfahrens. Bei dieser Auswertung wer-

den alle Aktionen, die einen Verbesserungsprozeß auslösen sollen protokolliert und anschließend auf ihren Erfolg überprüft.

5     **Mail-Listings:** Protokollierung sämtlicher eingehender Mails

**Alarmsystem:** In diesem Modul werden die Eingriffsgrenzen definiert. Hier soll vermieden werden, daß Seiten verweisen und nicht die nötige Beachtung im Evaluierungsprozeß finden.

10    **Zielvorgaben-Kontrolle:** Dieses Modul dient der Zielvorgaben-Kontrolle. Hier werden die angestrebten Zielvorgaben mit dem erreichten Ziel verglichen und von den normal zu erwartenden Steigerungsraten bereinigt. Das Ergebnis ist eine Über- bzw. Unter-

15    deckung der Zielvorgaben.

**e-commerce:** Dieses vergleichende Tool mißt das Verhältnis zwischen der Betrachtung von Informationen und den tatsächlichen Bestellungen. Dieses Verhältnis ist für weitere Optimierungsmaßnahmen im

20    Kommunikationsmix ein entscheidender Planungsfaktor.

**Pikanalyse:** Die Pikanalyse stellt alle Vorgänge dar, die außerhalb eines Trends liegen. Durch die Untersuchung von ungewöhnlichen Ausschlägen können

25    Rückschlüsse auf Werbemaßnahmen außerhalb des Webs getroffen werden.

Diese schon sehr ausführliche Dokumentation der historischen Daten, die natürlich von Interpretationsfehlern

30    bereinigt sind, ist für eine zielorientierte Planung nicht ausreichend. Hier kommt die vergleichende Analyse zum tragen, die zwischen relativem und absolutem Erfolg unterscheidet. Natürlich können die zu vergleichenden Zeiträume frei gewählt werden. Ebenso werden evtl. Ak-

tualisierungen, die die Auswertung verfälschen könnten, bereinigt dargestellt.

Das Verfahren zeichnet sich deshalb durch folgende zusätzliche Funktionalitäten aus:

**Vergleich beliebiger Zeiträume:** Beim Vergleich beliebiger Zeiträume werden alle Übersichtsdaten in den entsprechenden Zeitintervallen dargestellt.  
10 Die daraus resultierenden Ergebnisse zeigen den Erfolg verschiedenster Maßnahmen auf. Die Ergebnisse sind von den Aktualisierungen und damit verbundenen möglichen Fehlinterpretationen bereinigt.

**Steigerungsraten Kompensation:** Jede Internet-Domain  
15 erfährt eine natürliche Steigerungsrate. Dies liegt häufig an der steigenden Anzahl von Altkunden, die immer wieder die Domain besuchen bzw. an der steigenden Zahl von Internet- Usern. Diese natürliche Steigerungsrate, die nicht unbedingt die  
20 Erfolgssteigerung widerspiegelt, wird im Verfahren ermittelt und kompensiert. So wird der exakte Erfolg ermittelt, der von möglichen Fehlinterpretationen bereinigt ist.

**Analyse der Informationsbedürfnisse der User:** Die  
25 vergleichende Analyse erlaubt es, Änderungen der Informationsbedürfnisse der User festzustellen. Mit diesem Wissen kann der Content dem Userprofil angeglichen werden und so die Kundenzufriedenheit positiv beeinflusst werden. Dies ist insbesondere  
30 für das Binden von Altkunden von großer Bedeutung.

**Analyse der relativen Ausgewogenheit:** Zum Start einer Internet-Domain wird man feststellen, daß User nicht die gesamte Bandbreite der Informationen benötigen. Durchschnittlich dauert der erste Besuch

einer Page 4 Minuten. Aus diesem Grund ist es wichtig Benutzerprofile anzulegen, um daraus eine Ausgewogenheit der gesamten Domain zu erzeugen. Die symmetrische Verteilung von Visits auf contentführenden Seiten ist anzustreben.

**Analyse der Veränderungen von Rechnerfähigkeiten:**

Möchten Sie auch gerne wissen, ob Ihr Klientel die Cookies aktiviert hat oder nicht? Oder ob Sie endlich einmal eine andere Auflösung als Standard voraussetzen können? Mit der vergleichenden Analyse sehen Sie auf einen Blick die Veränderungen der Performance von Rechnern. Dies bietet genügend Raum rechtzeitig auf die veränderten Bedingungen reagieren zu können.

**Suchmaschinen:** Das Eintragen in Suchmaschinen ist sehr wichtig. Oft ist dies die einzige Möglichkeit Informationen zu beschaffen. Es reicht nicht aus, Anzeigen mit www. Adresse zu schalten oder diese auf der Visitenkarte zu erwähnen. Ein Eintrag auf den vorderen Plätzen ist von immenser Bedeutung. Dazu müssen die Erfolge der Suchmaschinen und die Suchworte analysiert und die Veränderungen abgespeichert werden.

**Verlinkungen:** Wie viele neue Besucher Ihrer Seite resultieren aus einer Verlinkung? Ist der Werbeträger geeignet für eine Plazierung von Werbebotschaften? Diese und weitere Fragen werden mit diesem Modul beantwortet.

**Alt-/Neukunde:** Sehr interessant sind die Veränderungen des Neu-/Altkunden Verhältnisses. Ist es möglich Kunden dauerhaft durch einen besonderen Service an die Domain zu binden? Ein ausgewogenes Verhältnis zwischen Neu-/Alt-kunden ist von großer Wichtig-

keit. Es zeigt auf, wieviel Vertrauen Sie zu Ihrem Kunden via Internet aufbauen können.

**Verweilzeiten:** Wie verändern sich die Verweilzeiten besonders im Altkunden-Segment? Die Antwort auf diese Frage drückt die Entspanntheit und das Vertrauen in den Informationsgehalt aus. Ebenso sind die Veränderungen der Userpfade wichtig für eine kundennahe Gestaltung der Domain.

**Steigerungsraten im e-commerce:** Die Veränderungen zwischen Informationsbeschaffung und der tatsächlichen Bestellung von Dienstleistungen, Informationen oder Produkten sind wichtige Kennzahlen für die Absatzplanung. Die hieraus resultierenden Ergebnisse lassen sich direkt in den entsprechenden Planungswerkzeugen einsetzen.

**Vergleichende Pikanalyse:** Abweichungen im normalen Zugriffsverhalten, die z.B. durch Anzeigen oder andere Aktionen ausgelöst werden, können hier genau verfolgt werden. Als Beispiel dient hier eine Reihe von Anzeigen, die im 14tägigen Abstand unverändert in derselben Zeitung erscheint. Nach einer gewissen Laufzeit werden die Ausschläge im Internet immer geringer. Dies kann als eindeutiges Zeichen für das „Abnutzen“ des Inserates gewertet werden. Die Konsequenz: Die Anzeigen-Aktion pausieren lassen oder eine neue Anzeige konzipieren.

Nachdem das Analyseverfahren richtige und bereinigte Daten ermittelt hat und diese in eine vergleichende Analyse münden, ist die logische Konsequenz die Ermittlung und Berechnung der Trends. Diese Trendberechnung wird nach DIN ISO durchgeführt. Als Basis werden statistische Modelle der Qualitätssicherung angewendet. Die-

se Berechnungen sind im Einklang mit der DGQ und dem TÜV cert entstanden.

Beantwortet werden folgende Fragen :

- 5     -     Wie entwickeln sich die Seitenzugriffe?
- Wie ändert sich zukünftig das zeitbezogene User-  
verhalten?
- Wie entwickelt sich die Rechner Performance?
- Welchen Einfluß werden zukünftig Suchmaschinen und  
10    Links auf den Erfolg einer Domain haben?
- Wie wird sich zukünftig das Verhältnis Neu-  
/Altkunde verändern und welche Auswirkungen wird  
dies haben?
- Wie wird sich zukünftig das Verhältnis zwischen  
15    Informationsbedarf und Bestellungen im e-commerce  
Sektor ändern?

Ferner benötigt der Webmaster einen Referenzpunkt, um  
seine eigene Position bestimmen zu können. Es stellt  
20    sich z.B. die Frage: "Ist das Verhältnis von 36:1 von  
Informationsbeschaffung zu tatsächlicher Bestellung po-  
sitiv oder negativ zu bewerten ? Diese Kalibrierung er-  
folgt durch eine externe Erfahrungsdatenbank.

In der externen Erfahrungsdatenbank wird das Know-how  
25    von vielen Domains komprimiert dargestellt. Sämtliche  
Daten sind von Erkennungsmerkmalen der Firmen berei-  
nigt. Diese Daten werden den Usern des Verfahrens on-  
line zur Verfügung gestellt. Hier werden Informationen  
über Trends im Internet, Erfahrungen im Bereich e-  
30    commerce, typische User-Profile sowie Informationen  
über das Userverhalten gespeichert sein. Aus diesen Er-  
fahrungen resultieren diverse Optimierungsmöglichkeiten  
für die vorhandene Internet Domain.

Zur Optimierung der Domain stehen diverse Tools zur Verfügung, die sich in drei generelle Kategorien einteilen lassen :

Kategorie 1 beinhaltet statische Module. Diese Module  
5 optimieren den Code der Domain in zweierlei Hinsicht. Zum einen werden die Module zur exakten Analyse integriert. Zum anderen werden mögliche Fehler kompensiert und korrigiert. Es ist ein „Übersetzer“ integriert, der die technische Information in eine anschauliche, dem  
10 Content entsprechende Darstellungsform transferiert. In diesem Abschnitt werden die Parameter und Zielvorgaben gesetzt, Fragen nach Eingriffsgrenzen gestellt und so das Verfahren initialisiert. Die in der Kategorie 1 befindlichen statischen Module sind:

- 15 - Codeoptimierung
- Frame-Korrektur
- Proxy-Optimierung
- Struktur-Konverter
- allgemeine Korrekturmodule
- 20 - Parameter Modul

Kategorie 2 beinhaltet dynamische Module, die Informationen während des Betriebes der Internet-Domain liefern. Zu diesen Modulen zählen die implementierten Statistik-Module und ein sogenannter Error-Messenger. Dieser  
25 schickt bei Auftritt eines Fehlers auf dem Server ein E-Mail an den Webmaster mit dem Vermerk, welcher Fehler beim User entstanden ist.

Der User erhält im Gegenzug die Meldung, daß dieser  
30 Fehler schnellstmöglich beseitigt wird. Diese Maßnahme ist ein entscheidender Schritt in Richtung Kundenorientierung.

Kategorie 3 beinhaltet dynamisch-administrative Module. Der Suchwort-Optimierer analysiert aus den Contentsei-

ten die Wörter, die eine hohe Frequenz aufweisen und erarbeitet daraus eine Vorschlagsliste. Diese Vorschlagsliste wird vom Webmaster angenommen bzw. verbessert. Diese Wörter werden anschließend in das Modul Meta-Optimierung einprogrammiert. Das Resultat ist ein Metatag, der allen Anforderungen der Suchmaschinen entspricht und den größtmöglichen Erfolg garantiert. Die Veränderungen werden im Suchmaschinen-Eintragsmodul gespeichert. Dieses Modul nimmt automatisch den Kontakt zu den gewählten Suchmaschinen auf und veranlaßt die Suchrobots zu einem Besuch der aktualisierten Domain. Durch diese Maßnahme ist es erstmals möglich diese Aufgabe teilweise zu automatisieren. So ist es möglich, das Ranking wesentlich zu verbessern und eine der oberen Positionen permanent beizubehalten. Weitere Module generieren Vorschlagslisten für sämtliche Aktionen, die offline betrieben werden müssen. Zu diesen zählt z.B. ein Vorschlagswesen für Anzeigenschaltungen.

Der nun innenliegende Regelkreis ist geschlossen: Messung, Vergleich, Trendberechnung, Optimierung, Übertragung und erneute Erfolgsmessung. Dieses Verfahren ermöglicht es Schritt für Schritt einen kontinuierlichen Verbesserungsprozeß einzuleiten, welcher von der DIN ISO 9000 gefordert wird.

Ein weitere Bestandteil des Verfahrens ist ein Lernprogramm für effektives Webcontrolling und ein Configurations-Tool. Dieses Tool vermittelt multimedial sämtliches Know-how, welches für den sukzessiven Aufbau und die Vermarktung von Content im Internet relevant ist. Hier können folgende Informationen abgerufen werden:

Was kann analysiert werden?

Wie sicher sind die Aussagen und die Daten?



Welche Datenschutzgesetze müssen beachtet werden?  
Welche Punkte müssen besondere Beachtung finden?  
Wie kann ein kontinuierlicher Verbesserungsprozeß erreicht werden?  
5 Wie wird ein effektives Webcontrolling am bestens strukturiert?

Im Modul „Configuration Tool“ werden dem Webmaster konkrete Fragen zu seinem Informationsbedürfnis gestellt.  
10 In diesem Bereich entsteht die Darstellungsform der Ergebnisse. Es erfolgt eine Nachbildung der Contentstruktur sowie eine Vorstellung des Verfahrens als Basis für das effektive Webcontrolling. Dieses Tool modifiziert und optimiert den Quellcode einer Domain. Weiterhin werden strukturelle Änderungen durchgeführt und  
15 bestimmte Merkmale hinzugefügt. Diese Modifikationen werden gesichert, so daß zu jedem Zeitpunkt der Originalzustand wieder hergestellt werden kann. Aus diesem Grund ist der zentrale Datenpool das Herzstück des Verfahrens.  
20

In einem weiteren Modul, dem zentralen Datenpool, wird eine Sicherung der Domain vor jeglicher Veränderung angelegt. Dort wird neben der alten Struktur der Daten  
25 die nun modifizierte Struktur angelegt. Die Parameter und die historischen Änderungen werden abgespeichert. So ist es möglich, bei der vergleichenden Analyse Weiterentwicklungen zu integrieren. Fehlinterpretationen sind so nahezu ausgeschlossen. Parallel zum klassischen  
30 Logfile erfolgt die Etablierung eines zweiten Logfiles. Dieser Logfile wird neben dem externen Logfile gespeichert. Sämtliche Daten sind nun unabhängig vom Provider zu führen; automatisches Downloaden von Logfiles ermöglicht das lückenlose Erfassen von Userdaten. Hier wer-

den sämtliche Analyse-Auswertungen gespeichert und in verständlicher Form im Analysebereich webmastergerecht dargestellt. Diese Informationen können auch über die Universalschnittstelle in bereits existierende Planungswerkzeuge integriert werden. Natürlich können auch diese Daten in individuellen Erweiterungsmodulen ihren Einsatz finden.

Die offene Plattform, die das Verfahren darstellt, erlaubt es weitere Werkzeuge zu kreieren. Diese Module werden individuell nach Kundenwunsch gefertigt. Zu diesen möglichen Produkten zählt das Mail-Controlling, das Kommunikationsmix-Controlling oder das Administrations-Controlling.

Der rasante Fortschritt in den modernen Kommunikationstechniken, wie z.B. ADSL, HDSL und power line communication ermöglicht ungeahnte multimediale Fähigkeiten im Internet. Die Inhalte werden stärker auf zielgruppenspezifische Aspekte ausgerichtet. Interaktivität fördert den Dialog mit den einzelnen Kundensegmenten. Somit können Service- und Dienstleistungsmerkmale besser hervorgehoben werden. Internet ermöglicht eine kostengünstige, kontinuierliche Kommunikationskette zum Kunden und bildet eine breitgefächerte Informationsbasis ohne unmittelbaren Kaufdruck. Die Berechnung des Erfolges bzw. return on investment (ROI) einer Internetseite ist entscheidend für die strategische und inhaltliche Weiterentwicklung des Content, der Kundenansprache und der Kundennähe.

Ziel des Verfahrens ist - im Sinne eines hochwertigen Reporting Systems - die Erhöhung der Interaktivität zum Kunden. Die Herausstellung von Produkt- bzw. Service-

vorteilen sowie die Optimierung der Logistik und des Kostenmanagements sind ebenfalls von hoher Relevanz.

Einfache Web-Site-Statistiken lassen sich durch Standard-Analysen erstellen. Diese bleiben jedoch generell unspezifisch, da sie keine Identifikation des Users und der User-Pfade sowie keine Analyse der Vertiefungsgrade der Abfrage, der Pausen, der Verweildauer etc. ermöglichen. Attraktive Inhalte und Kommunikationsinstrumente lassen sich jedoch nur durch eine personifizierte Analyse der Benutzerbedürfnisse realisieren. Daher ist die Zielgruppenanalyse, also die Unterteilung nach typischen Verhaltensmerkmalen, von immenser Bedeutung.

Um neue, innovative Dialog-Konzepte zu generieren, reicht die Standard-Analyse nicht aus. Nur eine Kombination verschiedener Methoden ermöglicht ein idealtypisches Dialog-Konzept und gewährleistet, daß alle wesentlichen Einflußfaktoren einer zukunftsweisenden und qualifizierten Kundenorientierung Berücksichtigung finden.

Der so zu realisierende Nutzen für den Anwender läßt sich wie folgt spezifizieren:

- Flexibilisierung und Optimierung des kundenorientierten Kommunikationsmix
- Dokumentation des Internet-Auftritts
- Erfolgsmessung und Offenlegung von Stärken und potentiellen Herausforderungen der Kommunikation
- Etablierung integrierter Kommunikationskonzepte durch Print-Medien, TV, Hörfunk,
- Direct Mailing und elektronische Medien

- unmittelbare Ableitung von Markt- und Nachfrage-trends und sofortige Anpassung des Kommunikations-mix
- Senkung der Medienbelegungs-, Logistik- und Print-kosten
- optimierte Zielgruppenansprache.

Der Funktionalitätsumfang des Verfahrens eröffnet ihm Einsatzmöglichkeiten, die vom privater Anbieter über lokale und regionale Anbieter bis hin zu national/multinationalen Anbietern reichen.

Der private Anbieter stellt auf einem virtuellen We-bospace persönliche Informationen zur Verfügung. Wie der private Anbieter verfügt der lokale Anbieter häufig über keinen eigenen Server. Sein Anforderungsprofil entspricht somit weitestgehend dem des privaten Anbie-ters. Beide werden mit nur einigen der beschriebenen Module auskommen, um ihre Anforderungen zu realisieren. Der regionale Anbieter stellt ein weitaus umfangreiche-res Informationsangebot zur Verfügung, da er zumeist über mehrere Dependancen in einer Region verfügt. Es ist davon auszugehen, daß dieser Kunde über ausgereif-tere technische Möglichkeiten verfügt, um verifizierte Daten über das User-Verhalten zu erhalten. Er ist somit potentieller Anwender des Verfahrens im vollen Umfang.

Die national/multinationalen Unternehmen verfügen zu-meist über einen eigenen Server und sind somit in der Lage, das gesamte Internet-Spektrum auszunutzen. Sie sind potentielle Kunden des gesamten Verfahrens ein-schließlich der Trainings-Tools. Je nach Unternehmens-größe empfiehlt sich außerdem die individuelle Program-mierung von Schnittstellen zu anderen handelsüblichen Warenwirtschaftsprogrammen und Planungswerkzeugen.

Das Dienstleistungskonzept beginnt mit der Ist-Analyse des Internetauftrittes des Kunden. Hierbei wird eine Stärken-/Schwächen-Analyse vorgenommen, die kundenseitige Verbesserungspotentiale aufzeigt.

5

Da nicht alle Features des Controlling Tools automatisiert werden können, werden einige Tools manuell entsprechend dem Anforderungsprofil des Kunden adaptiert.

10

Hieraus resultieren für den Kunden diverse Vorteile. So wird durch das "Outsourcen" von Software-Spezialisten jederzeit ein qualitativ hochwertiger Support sichergestellt und gleichzeitig eine Kostenreduktion im Bereich des Personalwesens des Kunden realisiert. So ist Spezialisten Know-how "on demand" abrufbar oder alternativ eine vor-Ort-Betreuung durch einen lokalen Spezialisten - z.B. einen Franchisenehmer- möglich.

15

20 Ein optimales Dienstleistungskonzept beinhaltet folgende Komponenten:

- Ist-Analyse
- Workshop zur Determinierung von Verbesserungspotentialen
- 25 - Festlegung des Soll-Zustandes
- Zieldefinition : kurzfristig - langfristig
- Umsetzung mit dem erfindungsgemäßen Verfahren
- Kontrolle

25

30 Per Ist-Analyse wird der aktuelle Zustand der Domain determiniert und alle verwertbaren Informationen aus dem Logfile entnommen. Anschließend erfolgt eine Potentialermittlung in einem Workshop in Kooperation mit dem

Kunden. Ferner werden hier der Soll-Zustand sowie die kurz- und langfristigen Ziele festgelegt.

5 Durch die Implementierung der einzelnen Module des Verfahrens, der Meta-Optimierung und der strukturellen Optimierung werden erste fehlerbereinigte Analysen des User-Verhaltens ermöglicht. Anhand historischer Daten erfolgt die Durchführung von Trendberechnungen, die als Basis für die weitere Planung dienen. Anschließend erfolgt eine kundengerechte Aufbereitung der Daten und  
10 daraus resultierend Optimierungsvorschläge.

Diese Ergebnisse werden mit den Daten der externen Datenbank sowie mit den Daten des durch das Verfahren angelegten Datenpools abgeglichen und evaluiert. Nach  
15 Präsentation dieser Informationen erfolgt ein erneuter Abgleich der kurz- und langfristigen Ziele und der daraus resultierenden Maßnahmen in Zusammenarbeit mit dem Kunden.

In Übereinstimmung mit diesen Zielen erfolgt nun die  
20 Content- und Strukturoptimierung sowie die Entwicklung der Corporate Identity und die Optimierung des Workflows. Hiermit schließt sich der Regelkreis, der in beliebigen Intervallen von 1 - 3 Monaten wiederholt wird.

25

Das Verfahren umfaßt folgende spezielle Dienstleistungen:

- Optimierung des Kommunikationsmixes durch gezielte Werbewirksamkeitsanalysen
- 30 - Etablierung von Sponsor-Konzepten durch exakte Analyse von Besucherzahlen und Zielgruppen
- Online-Werbung incl. Erfolgsmessung und Kosten-/Nutzen-Analyse

- Unterstützung von "Star-Allianzen" durch Schnittstellen-Programmierung zu Warenwirtschaftssystemen und Etablierung von Logistik-Konzepten
- Optimierung des Workflows von Bestellung bis zur  
5 Lieferung und Fakturierung in den jeweiligen Partnerunternehmen

Folgende Charakteristika unterstreichen die Einzigartigkeit des Dienstleistungs-konzepts für den Kunden :

- 10 - durch Outsourcing ressourcenschonende Personal- und Kostenstruktur
- innovative, "state-of-the-art"-Lösungen durch permanent geschultes Fachpersonal
- schnelle, messbare Erfolge durch zeitnahes Eta-  
15 blieren von Konzepten
- unabhängige und wertfreie Analyse des Userverhaltens durch qualitätsgesicherte Maßnahmen
- eindeutige Präsentation der Ergebnisse für die Entscheidungsträger
- 20 - unternehmensunabhängige Auswertung der Analyseergebnisse als "Stabsfunktion"
- Investitionsminimierung und -schutz durch sukzessive Investitionstätigkeit mit paralleler Erfolgsmessung
- 25 - Netzwerke durch Kooperation mit Marktführern und Know-how-Trägern zur Lösung individueller Herausforderungen

Die Erfindung ist nicht beschränkt auf die hier dargestellten Ausführungsbeispiele. Vielmehr ist es möglich,  
30 durch Kombination und Modifikation der genannten Mittel und Merkmale weitere Ausführungsvarianten zu realisieren, ohne den Rahmen der Erfindung zu verlassen.

### Patentansprüche

1. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz,  
5 dadurch gekennzeichnet, daß
  - nach Implementierung des Verfahrens der Quellcode der Domain optimiert wird,
  - von mindestens einem Servercomputer Benutzerakti-  
10 vitäten erfaßt und protokolliert werden,
  - die erfaßten Aktivitäten von Interpretationsfehlern bereinigt werden,
  - 15 - die nach der Bereinigung vorliegenden Daten analysiert und mit Daten aus einer Erfahrungsdatenbank verglichen werden und
  - 20 - die Ergebnisse des Vergleiches zur Optimierung eingesetzt werden, indem durch eine Rückführung eine aktive Einflußnahme auf den Servercomputer erfolgt und hierdurch ein geschlossener Regelkreis realisiert wird.
- 25 2. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach Anspruch 1,  
dadurch gekennzeichnet, daß  
30 die Benutzeraktivitäten in einem Logfile protokolliert und/oder in einer zweiten Datenbank gespeichert werden.



3. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach Anspruch 1 oder 2,  
dadurch gekennzeichnet, daß
- 5 im Logfile und/oder der zweiten Datenbank
- Parameter des Controllings und/oder
  - Daten über die Struktur der HTML-Dokumente und/oder
  - die Ist-Analyse und/oder

10 - die historischen Daten und/oder

  - die Trend-Analyse
- gespeichert werden.
4. Verfahren zur Analyse des Benutzerverhaltens in
- 15 Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 3,
- dadurch gekennzeichnet, daß
- das Verfahren eine Universalschnittstelle zur Verfügung stellt.
- 20
5. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 4,
- 25 dadurch gekennzeichnet, daß
- die protokollierten Benutzeraktivitäten über einen Zeitraum von mehreren Monaten gespeichert werden.
- 30 6. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach Anspruch 1,

dadurch gekennzeichnet, daß

die Erfahrungsdatenbank Informationen enthält

- über alle HTML-Dokumente, die, versehen mit Attributen, in Zuordnungs-Tabellen aufgelistet sind, und/oder
- über Parameter des Controlling.

7. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 6,

dadurch gekennzeichnet, daß

die den HTML-Dokumenten zugeordneten Attribute die HTML-Dokumente als

- technische Dokumente,
  - Dokumente zur Navigation,
  - contentführende Dokumente oder
  - werbeführende Dokumente
- kennzeichnen.

8. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 7,

dadurch gekennzeichnet, daß

die durch den Aufbau der Website und/oder die durch den Einsatz von

- Proxy-Cacheservern und/oder
- Lokalen Cachespeichern und/oder
- Routern und/oder
- Firewalls und/oder
- dynamisch vergebenen IP-Adressen

hervorgerufenen Interpretationsfehler bereinigt werden.

9. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß
- 5 jede Anfrage an eine dynamisch erzeugte Seite einschließlich der generierten Inhalte protokolliert wird.
10. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen nd zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß
- 10 jede Anfrage an eine Website zu einem Eintrag in das Logfile führt, indem eine Anfrage mittels eines CGI-Images immer dem Server mitgeteilt wird.
- 15 11. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 10, dadurch gekennzeichnet, daß
- 20 die Bereinigung der Interpretationsfehler mit Hilfe von Zuordnungstabellen erfolgt.
12. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 11, dadurch gekennzeichnet, daß
- 25 - die automatische Behebung von Programmierfehlern in den HTML-Dokumenten und/oder

- die automatische Anpassung der HTML-Dokumente an die gerätetechnischen Voraussetzungen der Nutzer und/oder
  - die automatische Anpassung der Inhalte der Websites an das Nutzerverhalten und/oder die Anforderungen der Nutzer
- Bestandteil der Optimierung sind.

13. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 12,
- dadurch gekennzeichnet, daß
- die Optimierung durch
- Vergleich mit dem typischen Zielgruppenverhalten und/oder
  - nach Methoden des kundenbezogenen Qualitätsmanagements
- erfolgt.

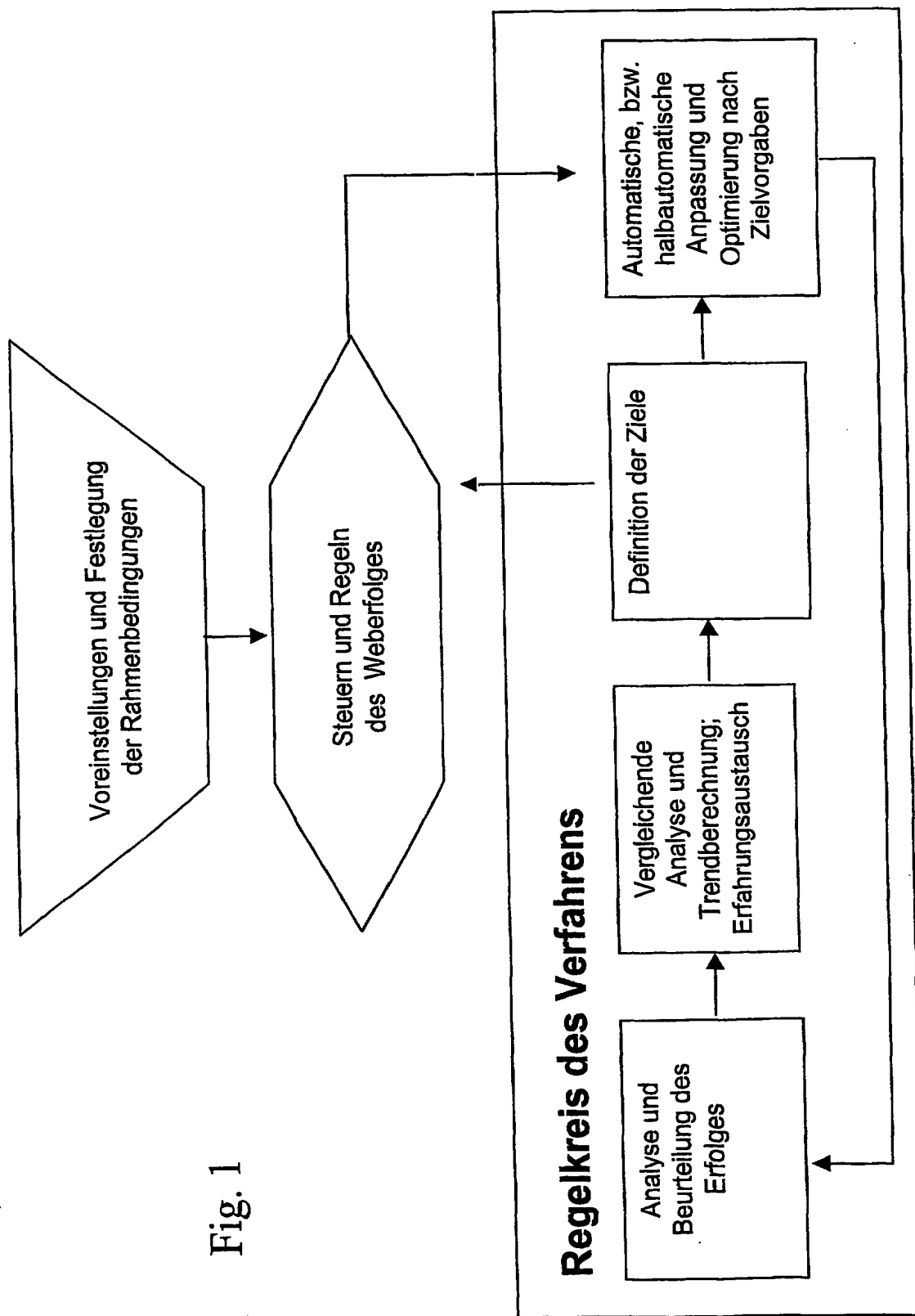
14. Verfahren zur Analyse des Benutzerverhaltens in Computernetzen zur Optimierung der Web-Präsenz nach einem der Ansprüche 1 bis 13,
- dadurch gekennzeichnet, daß
- eine Darstellung der Ergebnisse der Analyse nach Vorgabe individueller Parameter erfolgt.

15. Methode zur Optimierung der Web-Präsenz in Computernetzen,
- dadurch gekennzeichnet, daß
- das Benutzerverhalten erfasst und analysiert wird, wobei fehlerhafte Einflußfaktoren eliminiert werden

und eine Optimierung der Web-Präsenz mit dem Ziel der Erhöhung der betriebswirtschaftlichen Wirksamkeit der Web-Präsenz erfolgt derart, daß aus den erfassten, bereinigten Benutzeraktivitäten und Aussagen aus einer Erfahrungsdatenbank ein Regelwert zur Realisierung eines geschlossenen Regelkreises gebildet wird, welcher Einfluß auf die Gestaltung der Web-Präsenz ausübt.

**THIS PAGE BLANK (USPTO)**

1/7



**THIS PAGE BLANK (USPTO)**



2/7

# R.T.F.I. Leistungsspektrum im Überblick

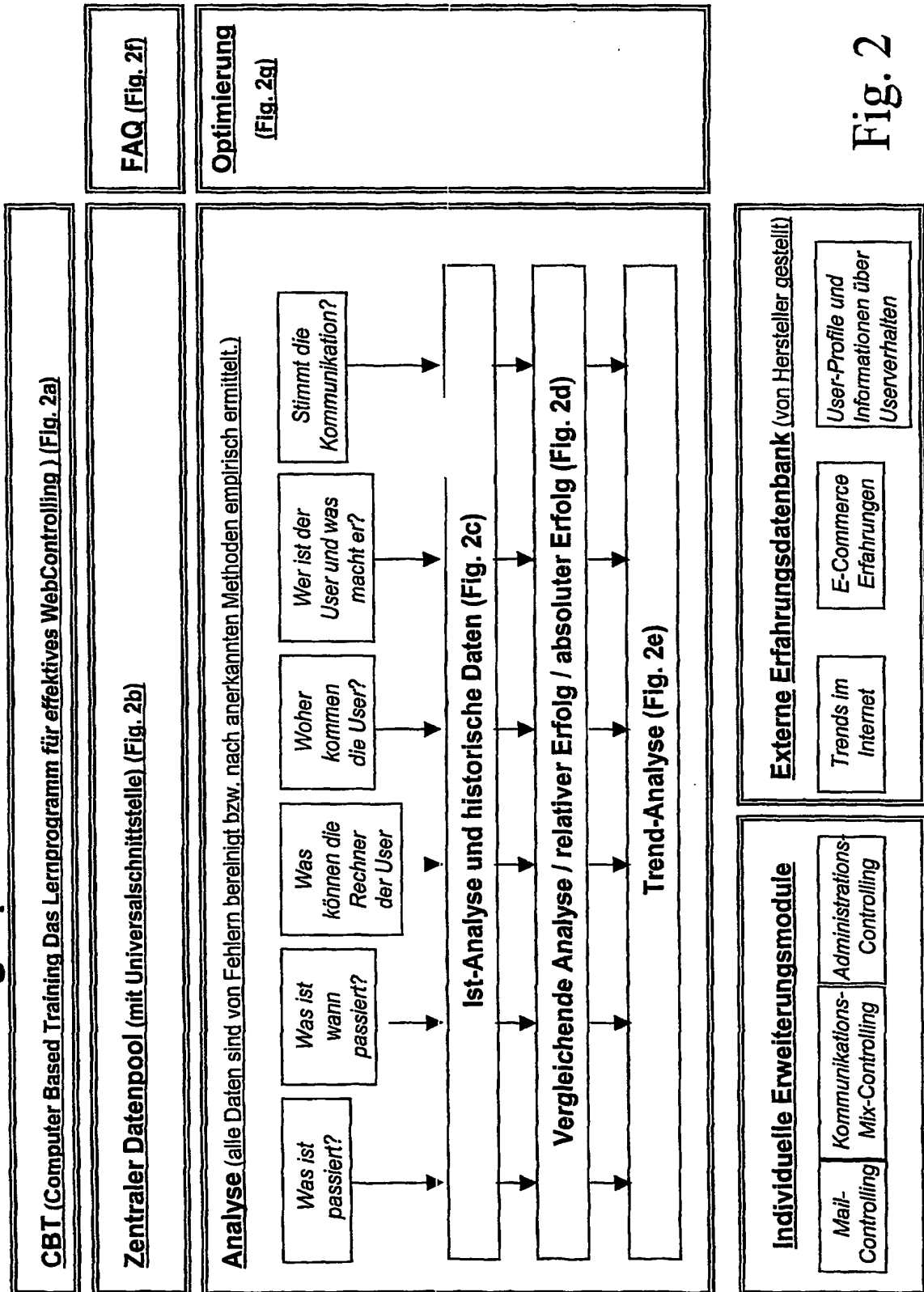
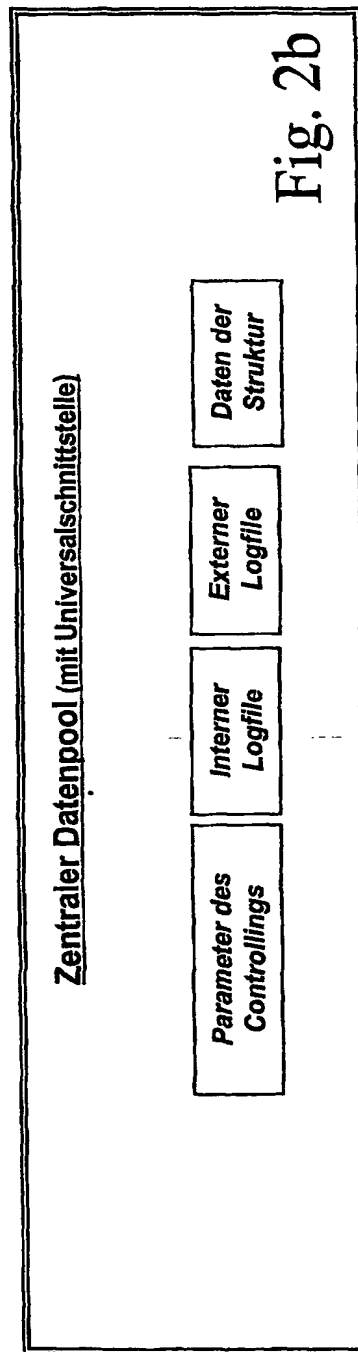
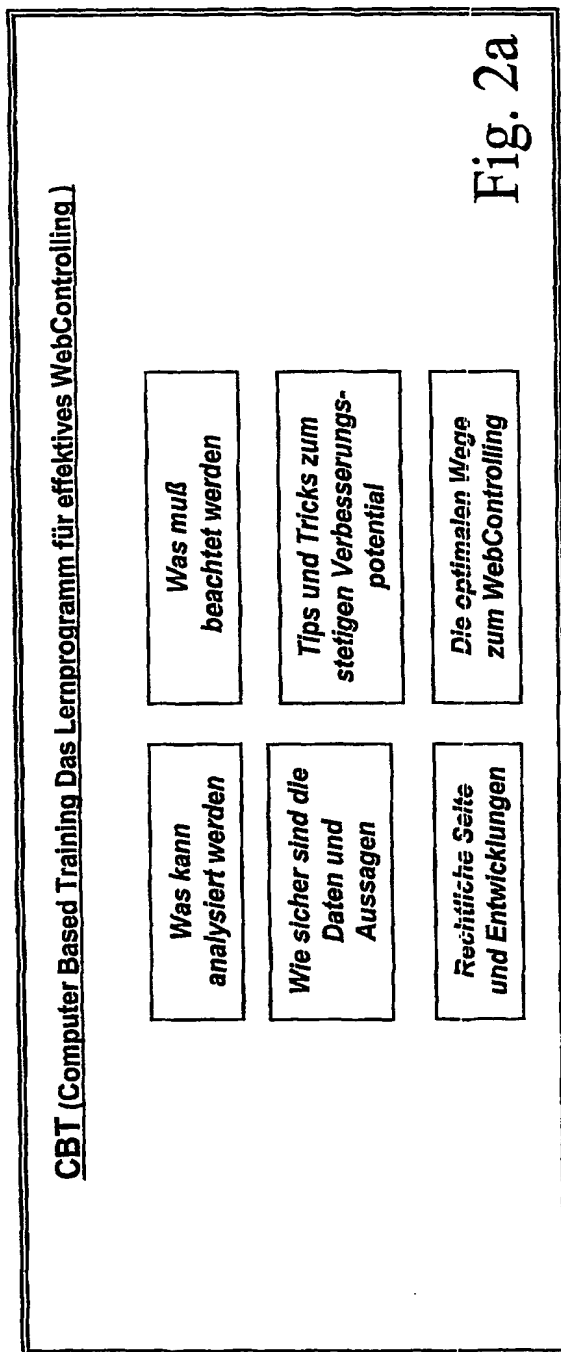


Fig. 2

**THIS PAGE BLANK (USPTO)**

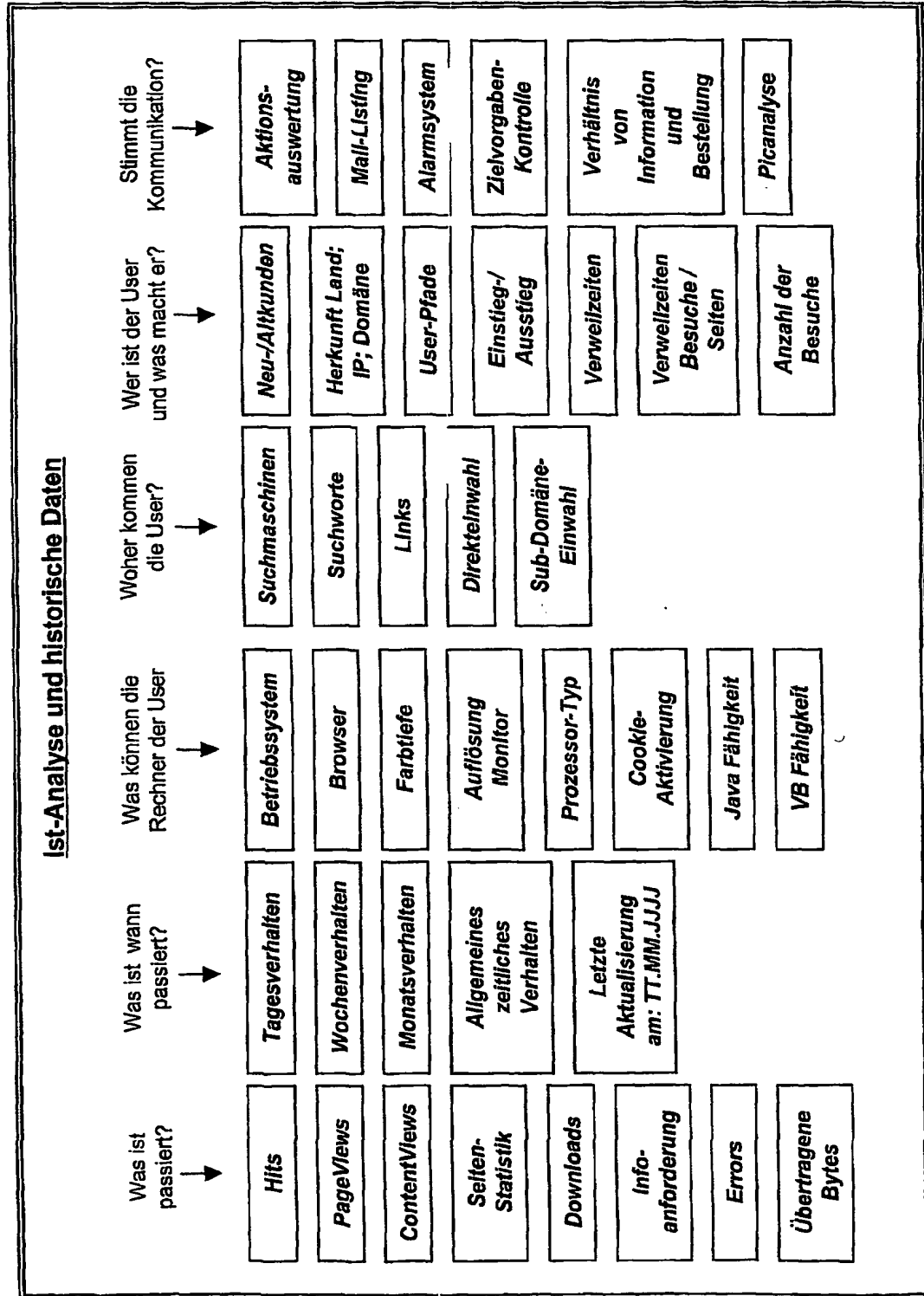
# R.T.F.I. Leistungsspektrum im Überblick



**THIS PAGE BLANK (USPTO)**

4/7

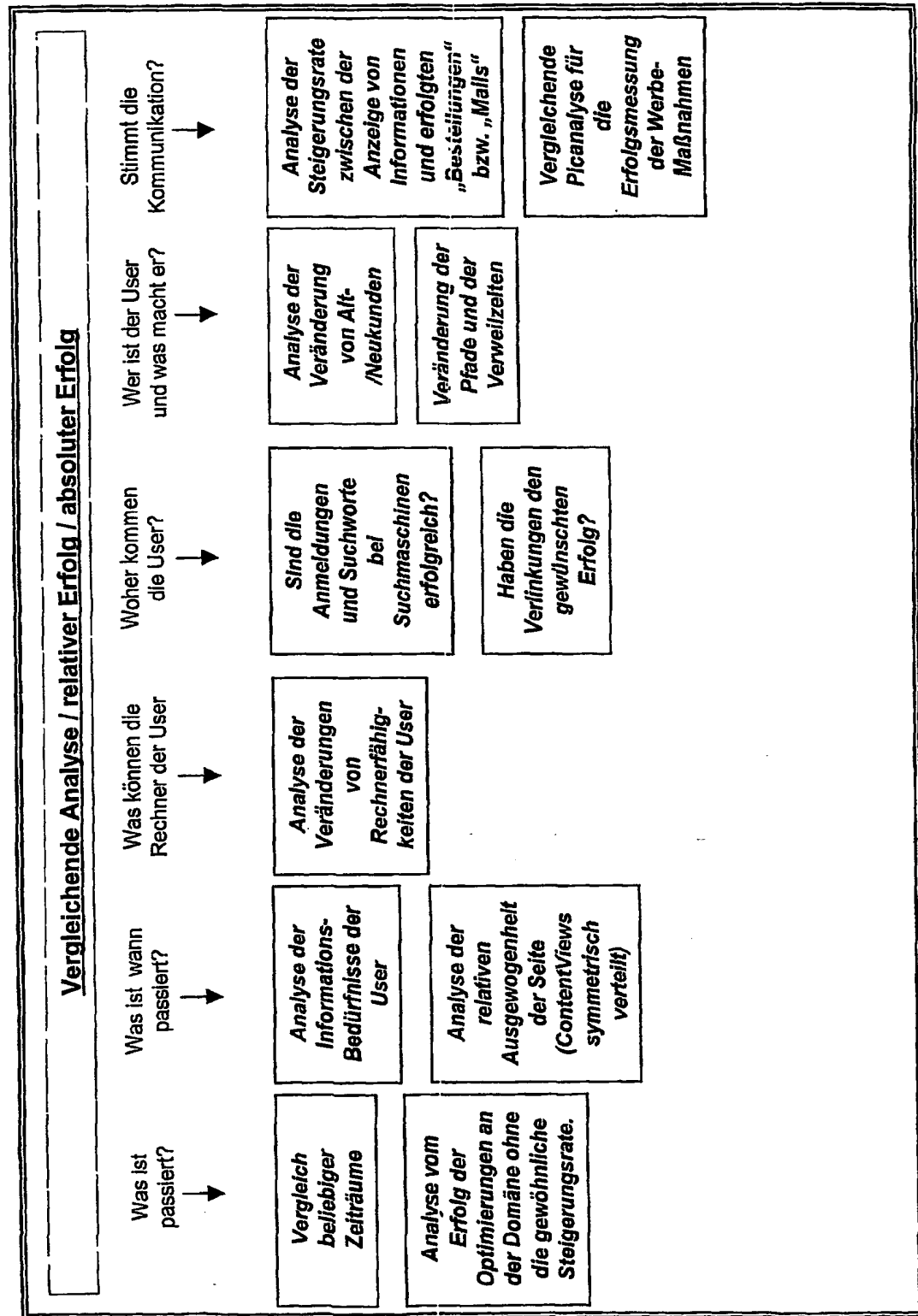
**R.T.F.I. Leistungsspektrum im Überblick** Fig. 2c



**THIS PAGE BLANK (USPTO)**

5/7

**R.T.F.I. Leistungsspektrum im Überblick** Fig. 2d



**THIS PAGE BLANK (USPTO)**



R.T.F.I. Leistungsspektrum im Überblick Fig. 2e

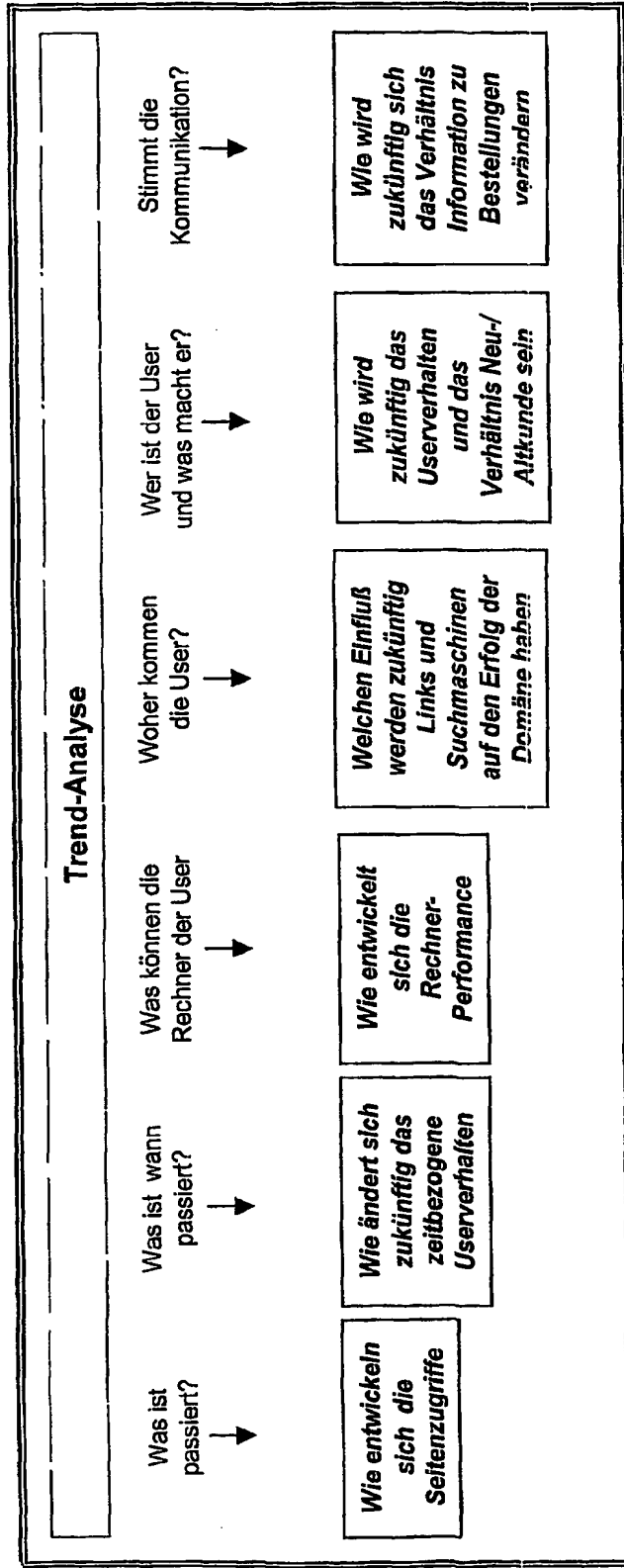
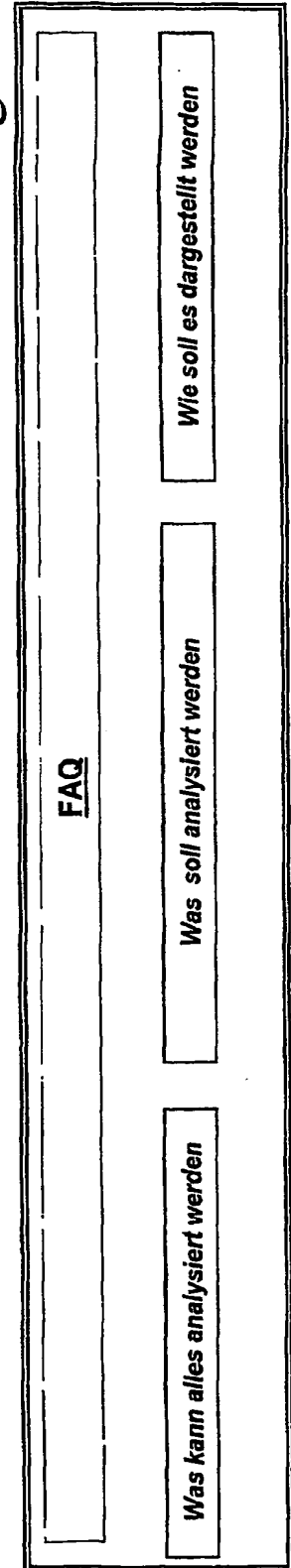
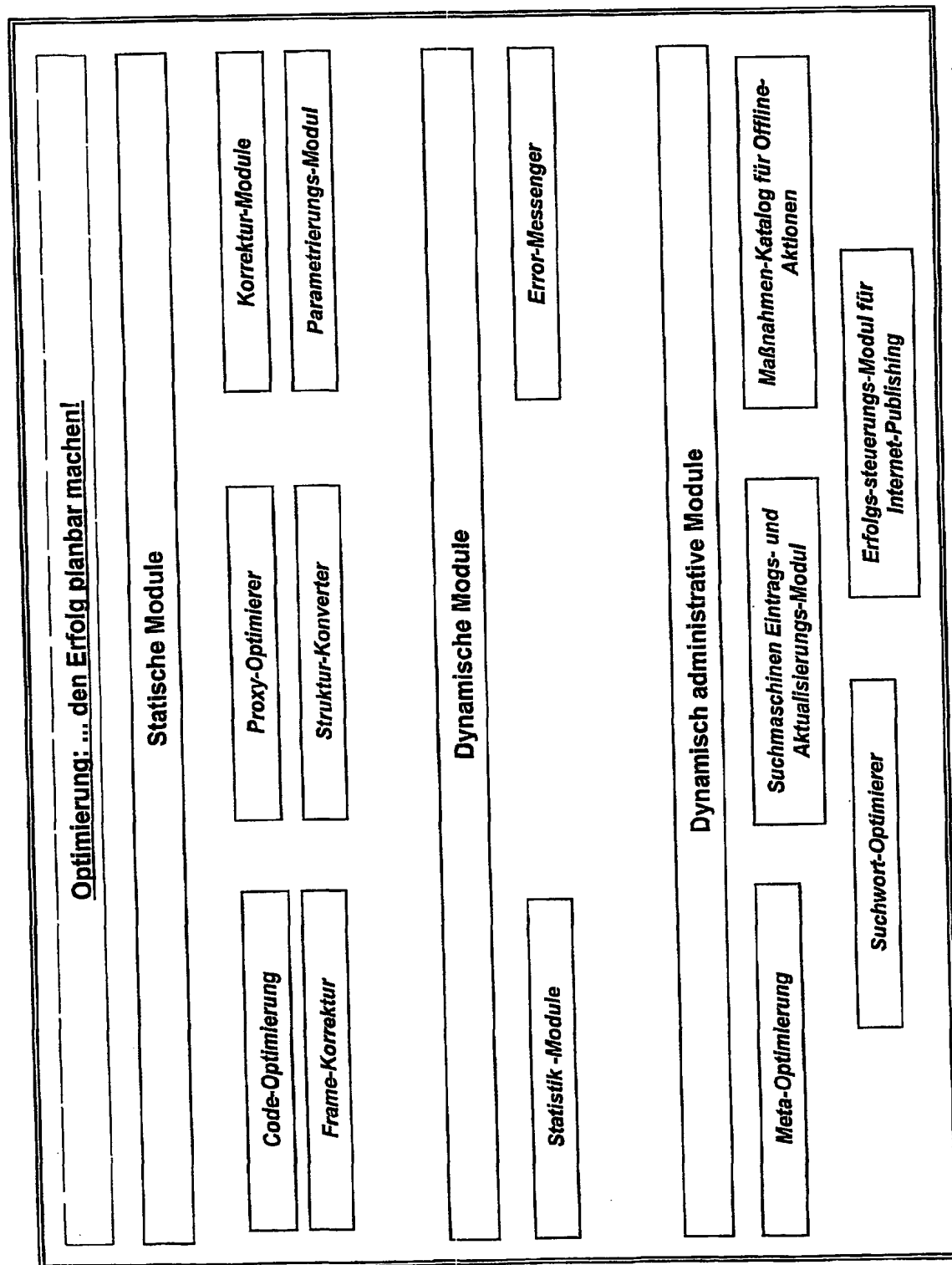


Fig. 2f



**THIS PAGE BLANK (USPTO)**

R.T.F.I. Leistungsspektrum im Überblick Fig. 2f



**THIS PAGE BLANK (USPTO)**

## INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 00/10136

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
IPK 7 G06F17/60

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 7 G06F

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, INSPEC, PAJ, COMPENDEX

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 5 892 917 A (MYERSON TERRY J) 6. April 1999 (1999-04-06) Zusammenfassung Spalte 2, Zeile 49 -Spalte 3, Zeile 37	1-15
X	COOLEY ET AL: "Grouping Web page references into transactions for mining World Wide Web browsing patterns" PROCEEDINGS OF THE KNOWLEDGE AND DATA ENGINEERING WORKSHOP, 1997, Seiten 2-9, XP002195450 Zusammenfassung Seite 6, linke Spalte, Absatz 2 -Seite 9, linke Spalte, Absatz 1	1-15
A	US 5 960 409 A (WEXLER D D) 28. September 1999 (1999-09-28) das ganze Dokument	1-15



Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen



Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

- \*A\* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- \*E\* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- \*L\* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- \*O\* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- \*P\* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

\*T\* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

\*X\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

\*Y\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

\*Z\* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

9. April 2002

Absendedatum des internationalen Recherchenberichts

23/04/2002

Name und Postanschrift der internationalen Recherchenbehörde  
Europäisches Patentamt, P.B. 5818 Patentaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Bevollmächtigter Bediensteter

Skulikaris, I

## INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 00/10136

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	PITKOW J: "In search of Reliable Usage Data on the WWW" PROCEEDINGS OF THE 6TH INTERNATIONAL WWW CONFERENCE, 1997, Seiten 451-463, XP002195451 Santa Clara, CA das ganze Dokument	1-15
A	MAH B A : "An empirical model of HTTP network traffic" PROCEEDINGS OF THE INFOCOM 97, 16TH ANNUAL JOINT CONFERENCE OF THE IEEE COMPUTER AND COMMUNICATION SOCIETIES, Bd. 2, 1997, Seiten 592-600, XP002195452 das ganze Dokument	1-15
A	FELCIANO R M ET AL: "Lamprey: Tracking Users on the World Wide Web" 1996 AMIA ANNUAL FALL SYMPOSIUM, 1996, XP002195453 Washington, D.C. das ganze Dokument	1-15

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 00/10136

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 5892917	A	06-04-1999	KEINE
US 5960409	A	28-09-1999	KEINE

**THIS PAGE BLANK (USPTO)**



## PARENT COOPERATION TREATY

PCT

## NOTIFICATION RELATING TO PRIORITY CLAIM

(PCT Rules 26bis.1 and 26bis.2 and  
Administrative Instructions, Sections 402 and 409)

From the INTERNATIONAL BUREAU

To:

HENGELHAUPT, Jürgen, D.  
Gulde Hengelhaupt Ziebig  
Schützenstrasse 15 - 17  
10117 Berlin  
ALLEMAGNE

Date of mailing (day/month/year) 15 May 2001 (15.05.01)	
Applicant's or agent's file reference P80299PC-HH	<b>IMPORTANT NOTIFICATION</b>
International application No. PCT/EP00/10136	International filing date (day/month/year) 13 October 2000 (13.10.00)
Applicant SCHLAFKE, Joachim	

The applicant is hereby **notified** of the following in respect of the priority claim(s) made in the international application.

1. ☒ **Correction of priority claim.** In accordance with the applicant's notice received on: 17 April 2001 (17.04.01), the following priority claim has been corrected to read as follows:

DE 15 October 1999 (15.10.99) 199 51 209.4

- ☐ even though the indication of the number of the earlier application is missing.  
☐ even though the following indication in the priority claim is not the same as the corresponding indication appearing in the priority document:

2. ☐ **Addition of priority claim.** In accordance with the applicant's notice received on: , the following priority claim has been added:

- ☐ even though the indication of the number of the earlier application is missing.  
☐ even though the following indication in the priority claim is not the same as the corresponding indication appearing in the priority document:

3. ☐ As a result of the correction and/or addition of (a) priority claim(s) under items 1 and/or 2, the (earliest) priority date is:

4. ☐ **Priority claim considered not to have been made.**

- ☐ The applicant failed to respond to the Invitation under Rule 26bis.2(a) (Form PCT/IB/316) within the prescribed time limit.  
☐ The applicant's notice was received after the expiration of the prescribed time limit under Rule 26bis.1(a).  
☐ The applicant's notice failed to correct the priority claim so as to comply with the requirements of Rule 4.10.

The applicant may, before the technical preparations for international publication have been completed and subject to the payment of a fee, request the International Bureau to publish, together with the international application, information concerning the priority claim. See Rule 26bis.2(c) and the PCT Applicant's Guide, Volume I, Annex B2(1B).

5. ☐ In case where **multiple priorities** have been claimed, the above item(s) relate to the following priority claim(s):

6. A copy of this notification has been sent to the receiving Office and

- ☒ to the International Searching Authority (where the international search report has not yet been issued).  
☒ the designated Offices (which have already been notified of the receipt of the record copy).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer  Dorothee Mülhausen
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

**THIS PAGE BLANK (USPTO)**

## PATENT COOPERATION TREATY

PCT

NOTIFICATION CONCERNING  
DOCUMENT TRANSMITTED

From the INTERNATIONAL BUREAU

To:

Commissioner  
US Department of Commerce  
United States Patent and Trademark  
Office, PCT  
2011 South Clark Place Room  
CP2/5C24  
Arlington, VA 22202  
ETATS-UNIS D'AMERIQUE  
in its capacity as designated Office

Date of mailing (day/month/year)  
15 May 2001 (15.05.01)

International application No.  
PCT/EP00/10136

International filing date (day/month/year)  
13 October 2000 (13.10.00)

Applicant

SCHLAFKE, Joachim

The International Bureau transmits herewith the following documents and number thereof:

\_\_\_\_\_ cop(ies) of priority document(s) (Rule 17.2(a))

The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Dorothee Mülhausen

Telephone No.: (41-22) 338.83.38

**THIS PAGE BLANK (USPTO)**

## PARENT COOPERATION TREATY

PCT

NOTIFICATION CONCERNING  
SUBMISSION OR TRANSMITTAL  
OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

From the INTERNATIONAL BUREAU

To:

HENGELHAUPT, Jürgen, D.  
Gulde Hengelhaupt Ziebig  
Schützenstrasse 15 - 17  
10117 Berlin  
ALLEMAGNE

Date of mailing (day/month/year) 15 May 2001 (15.05.01)	
Applicant's or agent's file reference P80299PC-HH	<b>IMPORTANT NOTIFICATION</b>
International application No. PCT/EP00/10136	International filing date (day/month/year) 13 October 2000 (13.10.00)
International publication date (day/month/year) 19 April 2001 (19.04.01)	Priority date (day/month/year) 15 October 1999 (15.10.99)
Applicant SCHLAFKE, Joachim	

1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).
2. This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.
3. An asterisk(\*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.
4. The letters "NR" appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, **the attention of the applicant is directed** to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.

<u>Priority date</u>	<u>Priority application No.</u>	<u>Country or regional Office or PCT receiving Office</u>	<u>Date of receipt of priority document</u>
15 Octo 1999 (15.10.99)	199 51 209.4	DE	21 Marc 2001 (21.03.01)

<p>The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland</p> <p>Facsimile No. (41-22) 740.14.35</p>	<p>Authorized officer</p> <p>Dorothee Mülhausen</p> <p>Telephone No. (41-22) 338.83.38</p>
---	--

**THIS PAGE BLANK (USPTO)**

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT  
AUF DEM GEBIET DES PATENTWESENS

PCT

10/8 68,110

INTERNATIONALER RECHERCHENBERICHT

(Artikel 18 sowie Regeln 43 und 44 PCT)

Aktenzeichen des Anmelders oder Anwalts <b>P80299PC-HH</b>	<b>WEITERES VORGEHEN</b> siehe Mitteilung über die Übermittlung des internationalen Recherchenberichts (Formblatt PCT/ISA/220) sowie, soweit zutreffend, nachstehender Punkt 5	
Internationales Aktenzeichen <b>PCT/EP 00/10136</b>	Internationales Anmeldedatum (Tag/Monat/Jahr) <b>13/10/2000</b>	(Frühestes) Prioritätsdatum (Tag/Monat/Jahr) <b>15/10/1999</b>
Anmelder <b>SCHLAFKE, Joachim</b>		

Dieser internationale Recherchenbericht wurde von der Internationalen Recherchenbehörde erstellt und wird dem Anmelder gemäß Artikel 18 übermittelt. Eine Kopie wird dem Internationalen Büro übermittelt.

Dieser internationale Recherchenbericht umfaßt insgesamt 3 Blätter.

☒ Darüber hinaus liegt ihm jeweils eine Kopie der in diesem Bericht genannten Unterlagen zum Stand der Technik bei.

1. Grundlage des Berichts

- a. Hinsichtlich der **Sprache** ist die internationale Recherche auf der Grundlage der internationalen Anmeldung in der Sprache durchgeführt worden, in der sie eingereicht wurde, sofern unter diesem Punkt nichts anderes angegeben ist.

☐ Die internationale Recherche ist auf der Grundlage einer bei der Behörde eingereichten Übersetzung der internationalen Anmeldung (Regel 23.1 b)) durchgeführt worden.

- b. Hinsichtlich der in der internationalen Anmeldung offenbarten **Nucleotid- und/oder Aminosäuresequenz** ist die internationale Recherche auf der Grundlage des Sequenzprotokolls durchgeführt worden, das

☐ in der internationalen Anmeldung in Schriftlicher Form enthalten ist.

☐ zusammen mit der internationalen Anmeldung in computerlesbarer Form eingereicht worden ist.

☐ bei der Behörde nachträglich in schriftlicher Form eingereicht worden ist.

☐ bei der Behörde nachträglich in computerlesbarer Form eingereicht worden ist.

☐ Die Erklärung, daß das nachträglich eingereichte schriftliche Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht, wurde vorgelegt.

☐ Die Erklärung, daß die in computerlesbarer Form erfaßten Informationen dem schriftlichen Sequenzprotokoll entsprechen, wurde vorgelegt.

2. ☐ Bestimmte Ansprüche haben sich als nicht recherchierbar erwiesen (siehe Feld II).

3. ☐ Mangelnde Einheitlichkeit der Erfindung (siehe Feld II).

4. Hinsichtlich der **Bezeichnung der Erfindung**

☒ wird der vom Anmelder eingereichte Wortlaut genehmigt.

☐ wurde der Wortlaut von der Behörde wie folgt festgesetzt:

5. Hinsichtlich der **Zusammenfassung**

☒ wird der vom Anmelder eingereichte Wortlaut genehmigt.

☐ wurde der Wortlaut nach Regel 38.2b) in der in Feld III angegebenen Fassung von der Behörde festgesetzt. Der Anmelder kann der Behörde innerhalb eines Monats nach dem Datum der Absendung dieses internationalen Recherchenberichts eine Stellungnahme vorlegen.

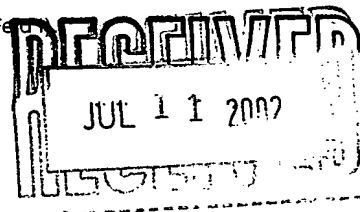
6. Folgende Abbildung der **Zeichnungen** ist mit der Zusammenfassung zu veröffentlichen: Abb. Nr. 1

☒ wie vom Anmelder vorgeschlagen

☐ weil der Anmelder selbst keine Abbildung vorgeschlagen hat.

☐ weil diese Abbildung die Erfindung besser kennzeichnet.

☐ keine der Abb.



**THIS PAGE BLANK (USPTO)**